nature

www.nature.com/nature

Vol 446 | Issue no. 7138 | 19 April 2007

Plus ça change?

The winner of France's election will find scientists willing to support the right kind of reforms.

In France's presidential election, the first round of which will be held on 22 April with a run-off two weeks later, research and innovation have emerged as significant campaign issues, crystallizing public concern about the nation's competitiveness and global standing.

This election is an important one for science. Presidents of France can make a real difference: Charles de Gaulle helped to create France's world-beating nuclear, aerospace and transport industries through a united national research effort in the 1950s. François Mitterrand invested heavily in research after his election in 1981 and oversaw a resurgence of French science.

Can France's next president fill these shoes? To help find out, *Nature* asked the three leading candidates — conservative Nicolas Sarkozy, socialist Ségolène Royal and a centrist, François Bayrou — to present their own visions for the future (see page 847).

All three pledge substantial spending on research to reverse the cuts imposed during Jacques Chirac's term of office. Increased funding is not enough, however. The next president must support much-needed, targeted reforms of the science system, which has resisted comprehensive reform efforts for more than two decades.

Under Chirac, successive conservative governments have unsuccessfully sought to impose ill-considered reform packages on scientists with little consultation. It hasn't helped that they have cut budgets at the same time. The French left has traditionally been more supportive of science, but has been unwilling to engage in the reforms that are needed to boost scientific performance.

The scientific community has already reached a broad consensus on the sort of reforms that are required (see page 850). A pragmatic restructuring of France's fragmented life-science research is a priority, for example. An upgrade of the science ministry from its current junior status, to put it on a par with agriculture, finance and foreign affairs, is another. Salary structures should be adjusted to enable French institutions to compete for international talent in areas such as information technology, and skilled immigrants should be welcomed, not discouraged.

Learning from the past

There is no real need for a vast national consultation exercise (a favoured French pursuit). France's next president, whoever it is, could use his or her post-election momentum to press forward with change. In doing so, the president should note the mistakes of the past: there has been a repetitive cycle whereby new governments announce sweeping reforms, scientists take to the streets to oppose them, and then stalemate sets in. France does not need a grandiose rearrangement of the deckchairs of its research administration; it needs pragmatic, well-targeted and sustained changes in areas such as recruitment policy and salary structures.

The scientific community, for its part, should do more to articulate reasonable concerns without falling into the trap of resisting all change, as some trade-union representatives have been inclined to do. Representatives of the community should reflect scientists' own hunger for steps that will nurture research excellence.

In our survey of the candidates' views, Sarkozy perhaps articulates the need for reform most clearly. He proposes the transformation of the research agencies — whose labs currently perform the majority of French research — into research councils that would fund labs within a powerful and autonomous university system.

Appealing as that sounds, it is essentially the same reform programme that Chirac and successive conservative governments supported but failed to implement. Many of Sarkozy's science advisers are familiar from earlier administrations. The campaign rhetoric of every conservative government over the past quarter-century has been to declare research a 'national priority', only to change its tune once elected. Chirac, for example, promised in 2002 that the "commitment to research must be historic", only to make harsh cuts that provoked historic street demonstrations.

Shift in emphasis

Bayrou is a relatively unknown quantity to the research community. In his responses to our questions, he avoids talk of wholesale change, promising instead that a cross-party, non-partisan consensus could be found on pragmatic improvements to the existing system. His comments strike the right tone, although it remains unclear how Bayrou's small centrist party could form a government.

Royal seems likely, at least, to keep her funding promises. She introduced science and education into the heart of her campaign, and her commitment to both seems genuine. The danger, as always with the left in France, is that she will baulk at any meaningful reform in the face of resistance from public-sector workers and the trade unions. Royal says she plans to replace France's aloof style of government with the pragmatic consensus-building associated with Scandinavia. Just what French research needs, perhaps — but easier said than done.

The realistic path forward for French science lies in a shift of emphasis and power away from the research organizations, such as the CNRS and INSERM, towards the universities. But that cannot occur overnight. France's universities need to be fixed, for a start. With a few exceptions, they are physically dilapidated, badly managed and host to nepotism. To transfer research to the universities in their current state would be a step backwards in terms of research excellence.

Researchers also accept the idea of a related move towards a dual system of government laboratories and competitive grants. The open question is how to do this, and how quickly. The National Research Agency (ANR), created under Chirac to distribute research funds on the basis of grant proposals, is a step in the right direction. But as a branch of the science ministry, it lacks the autonomy and authority of, say, the US National Science Foundation or Germany's DFG.

Priority should be given to reforms that will make the greatest impact. French science has great strengths, however: would-be reformers must be careful not to throw the baby out with the bathwater. ■

Small steps forward

The presidency of Olusegun Obasanjo, whatever its pitfalls, has been positive for Nigerian science.

The orderly transfer of power is an important test of any democracy. This Saturday, voters will go to the polls in Nigeria to select a successor to President Olusegun Obasanjo, the former general elected in 1999. Having served two terms, Obasanjo will return to academic life. But the election campaign has already been characterized by considerable violence, and it remains to be seen if its outcome can be regarded as fair.

Over the course of his presidency, Obasanjo has emerged as one of the most significant figures in contemporary African politics and has attempted to incorporate science into the continent's development. In recent months, he has given the go-ahead for reforms to Nigeria's science and technology governance, which had stagnated since independence in 1960. The reforms include creating a new National Council for Research and Development (to be chaired by his successor), and a number of individual research councils, which will provide funding on a competitive basis for research in specific areas such as agriculture and energy, as well as provision for independent quality assessments in the universities.

New research centres have also been established, including the African Institute for Science and Technology in Abuja. But perhaps his most ambitious plan was that of a major expansion, announced in June 2006, of the country's Petroleum Technology Development Fund, to set aside US$5 billion of oil revenues for research grants, education and infrastructure. The fund has been in existence for years, but the expansion could go some way towards protecting Nigerian researchers from the inevitable budget cuts that take place in oil-dependent nations when oil prices fall, or if political priorities change.

Unfortunately, the petroleum fund is now at the heart of a dispute between the departing president and the vice-president, Atiku Abubakar, who is standing for the presidency this time round. Each has accused the other of diverting money from the fund to non-scientific causes (see www.scidev.net/News/index.cfm?fuseaction=readNews&itemid=3334). So the jury remains out on whether the fund is ever going to materialize in its advertised form.

All of Nigeria's main opposition parties seem to be signed up to the Obasanjo agenda for science and development. The next president will inherit a range of instruments for promoting education and science — as well as a massive challenge in addressing poverty, disease and corruption. According to the United Nations Development Programme, one in two Nigerians lacks access to clean water, and life expectancy is less than 44, unmoved since 1970.

But at least Obasanjo recognized that Nigeria needs to invest in education and research in order to confront these challenges — an approach that is growing more popular across sub-Saharan Africa. That has provided a platform on which his successor can build. ■

Addicted to secrecy

Sealed drug documents should be opened up.

In the course of lengthy, expensive litigation over drug safety in the United States, opposing sets of lawyers often form unholy alliances. These cases are frequently settled before reaching court, and the two sides agree to confidentiality orders being placed on any drug-company documents to which they have been granted access. This suits the courts, which want parties to cases under their jurisdiction to gain access to as much information as possible; the drug companies, who want to keep most of the data confidential; and the plaintiffs' lawyers, who want to win a good deal for their clients. But it may not suit the wider public interest.

Every now and then, some of those sealed documents leak out. In a recent trial concerning the schizophrenia drug Zyprexa, for example, documents emerged that suggested to some parties that the manufacturer, Eli Lilly, had sought to play down some known side-effects of the drug (see page 838). But documents considered during pre-trial negotiations usually remain sealed by private agreement between the court and both sets of attorneys — even where there is a public-health interest in releasing some of the information they contain.

Some plaintiffs' attorneys have argued for the creation of ethical guidelines or legislation that would bar lawyers from reaching such agreements. Drug companies have countered that the agreements are helpful to the fair resolution of cases because they allow fairly untrammelled access to information relevant to each case. Without such agreements, defendants would doubtless resist the release of both personal and commercially sensitive information, slowing down cases and reducing their chances of just resolution.

Even so, a few states, such as Florida, have introduced anti-secrecy laws. But there are also other reform options. Courts could, for example, be allowed to take previous secrecy orders into consideration when setting damages. Drug companies might think twice about sealing data from clinical trials if they knew they could come out in a future trial and incur greater financial cost.

"Scientific bodies have an interest in ensuring that public-health data are released into the public domain."

Scientific bodies have an interest in ensuring that as many public-health data as possible are released into the public domain. The Institute of Medicine and the American Association for the Advancement of Science (which already has a joint committee with the American Bar Association to consider such matters) could help by fostering discussions between the plaintiffs' lawyers and the drug industry on how to move the issue forward.

The industry's reputation, as well as the public good, will benefit in the long run from arrangements that get pertinent information about drug safety into the public domain as quickly as possible. Global registers of clinical-trial results are probably the most pressing requirement here. But less court-imposed secrecy around public health would also be a positive step, in part by helping to ensure that companies are fully complying with such registers. ■

Small steps forward

The presidency of Olusegun Obasanjo, whatever its pitfalls, has been positive for Nigerian science.

The orderly transfer of power is an important test of any democracy. This Saturday, voters will go to the polls in Nigeria to select a successor to President Olusegun Obasanjo, the former general elected in 1999. Having served two terms, Obasanjo will return to academic life. But the election campaign has already been characterized by considerable violence, and it remains to be seen if its outcome can be regarded as fair.

Over the course of his presidency, Obasanjo has emerged as one of the most significant figures in contemporary African politics and has attempted to incorporate science into the continent's development. In recent months, he has given the go-ahead for reforms to Nigeria's science and technology governance, which had stagnated since independence in 1960. The reforms include creating a new National Council for Research and Development (to be chaired by his successor), and a number of individual research councils, which will provide funding on a competitive basis for research in specific areas such as agriculture and energy, as well as provision for independent quality assessments in the universities.

New research centres have also been established, including the African Institute for Science and Technology in Abuja. But perhaps his most ambitious plan was that of a major expansion, announced in June 2006, of the country's Petroleum Technology Development Fund, to set aside US$5 billion of oil revenues for research grants, education and infrastructure. The fund has been in existence for years, but the expansion could go some way towards protecting Nigerian researchers from the inevitable budget cuts that take place in oil-dependent nations when oil prices fall, or if political priorities change.

Unfortunately, the petroleum fund is now at the heart of a dispute between the departing president and the vice-president, Atiku Abubakar, who is standing for the presidency this time round. Each has accused the other of diverting money from the fund to non-scientific causes (see www.scidev.net/News/index.cfm?fuseaction=readNews&itemid=3334). So the jury remains out on whether the fund is ever going to materialize in its advertised form.

All of Nigeria's main opposition parties seem to be signed up to the Obasanjo agenda for science and development. The next president will inherit a range of instruments for promoting education and science — as well as a massive challenge in addressing poverty, disease and corruption. According to the United Nations Development Programme, one in two Nigerians lacks access to clean water, and life expectancy is less than 44, unmoved since 1970.

But at least Obasanjo recognized that Nigeria needs to invest in education and research in order to confront these challenges — an approach that is growing more popular across sub-Saharan Africa. That has provided a platform on which his successor can build. ■

Addicted to secrecy

Sealed drug documents should be opened up.

In the course of lengthy, expensive litigation over drug safety in the United States, opposing sets of lawyers often form unholy alliances. These cases are frequently settled before reaching court, and the two sides agree to confidentiality orders being placed on any drug-company documents to which they have been granted access. This suits the courts, which want parties to cases under their jurisdiction to gain access to as much information as possible; the drug companies, who want to keep most of the data confidential; and the plaintiffs' lawyers, who want to win a good deal for their clients. But it may not suit the wider public interest.

Every now and then, some of those sealed documents leak out. In a recent trial concerning the schizophrenia drug Zyprexa, for example, documents emerged that suggested to some parties that the manufacturer, Eli Lilly, had sought to play down some known side-effects of the drug (see page 838). But documents considered during pre-trial negotiations usually remain sealed by private agreement between the court and both sets of attorneys — even where there is a public-health interest in releasing some of the information they contain.

Some plaintiffs' attorneys have argued for the creation of ethical guidelines or legislation that would bar lawyers from reaching such agreements. Drug companies have countered that the agreements are helpful to the fair resolution of cases because they allow fairly untrammelled access to information relevant to each case. Without such agreements, defendants would doubtless resist the release of both personal and commercially sensitive information, slowing down cases and reducing their chances of just resolution.

Even so, a few states, such as Florida, have introduced anti-secrecy laws. But there are also other reform options. Courts could, for example, be allowed to take previous secrecy orders into consideration when setting damages. Drug companies might think twice about sealing data from clinical trials if they knew they could come out in a future trial and incur greater financial cost.

"Scientific bodies have an interest in ensuring that public-health data are released into the public domain."

Scientific bodies have an interest in ensuring that as many public-health data as possible are released into the public domain. The Institute of Medicine and the American Association for the Advancement of Science (which already has a joint committee with the American Bar Association to consider such matters) could help by fostering discussions between the plaintiffs' lawyers and the drug industry on how to move the issue forward.

The industry's reputation, as well as the public good, will benefit in the long run from arrangements that get pertinent information about drug safety into the public domain as quickly as possible. Global registers of clinical-trial results are probably the most pressing requirement here. But less court-imposed secrecy around public health would also be a positive step, in part by helping to ensure that companies are fully complying with such registers. ■

RESEARCH HIGHLIGHTS

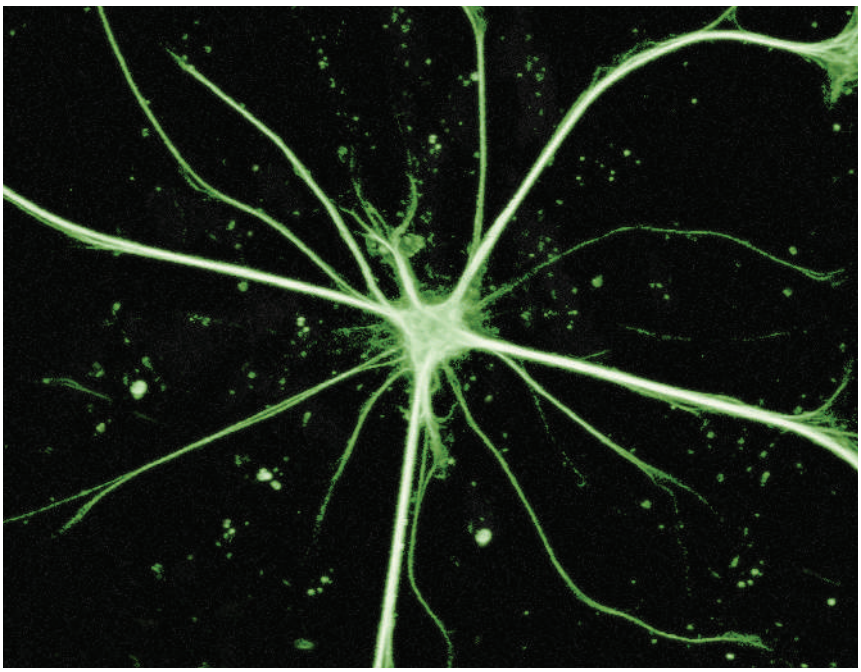
The poison next door

Nature Neurosci. doi:10.1038/nn1885 and doi:10.1038/nn1876 (2007)

The neurodegenerative disease amyotrophic lateral sclerosis (ALS) causes deterioration of muscle control by killing motor neurons. Now, two studies suggest that therapies should focus not on the motor neurons, but on neighbouring non-neuronal cells known as glia.

Roughly 2% of ALS cases are caused by a mutation in a gene known as superoxide dismutase-1 (*SOD1*). Serge Przedborski and his colleagues at Columbia University in New York found that a particular type of glial cell, an astrocyte (pictured), carrying the *SOD1* mutation releases an unknown factor that kills motor neurons.

Kevin Eggan and Tom Maniatis of Harvard University in Cambridge, Massachusetts, also found that mutant glial cells cause neurodegeneration. To perform their experiments, this team derived motor neurons with the SOD1 mutation from mouse embryonic stem cells. This provides a new system for rapidly screening candidate drugs.



SPL

GENETICS

Knockout round

Genome Res. doi:10.1101/gr.6080607 (2007)

Researchers in the Netherlands have developed an efficient approach to knocking out genes in the worm *Caenorhabditis elegans*, one of biology's favourite model organisms. Such knockout worms are used to study the disabled gene's function.

Edwin Cuppen and his colleagues at the Hubrecht Laboratory in Utrecht exposed worms to a chemical mutagen, then used high-throughput sequencing to look for mutations in 32 selected genes. They found, in more than 6,000 worms, mutations that inactivated 27 of the genes studied.

Their sequencing method is faster than existing techniques to identify knockouts, say the researchers, raising hopes that it could be extended to span all of the worm's nearly 20,000 genes.

TAXONOMY

Identity crisis

Proc. R. Soc. Lond. B doi:10.1098/rspb.2007.0248 (2007)

Many of the European medicinal leeches available commercially and marketed as *Hirudo medicinalis* are actually *Hirudo verbana*, according to a DNA study. *H. medicinalis* is approved by the US Food and Drug Administration (FDA) as a medical device, but *H. verbana* currently falls outside FDA approval.

Mark Siddall of the American Museum of Natural History in New York and his colleagues made the finding after analysing mitochondrial and nuclear DNA from 36 commercial and wild leeches from across Europe. They distinguish three species of European medicinal leech.

Research labs working with leeches — a popular model organism in neurobiology — will need to review their results if they find that their organisms have been misclassified.

ANIMAL BEHAVIOUR

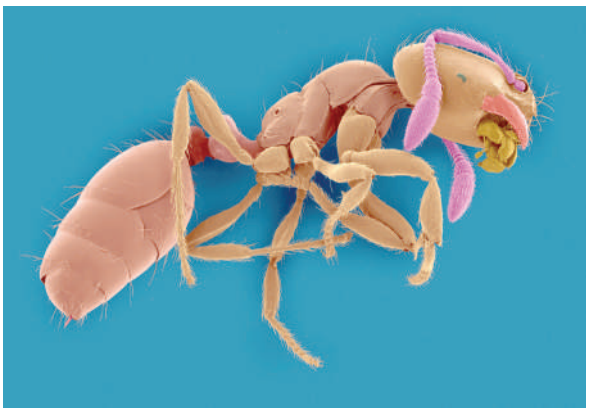
Handy legs

Naturwissenschaften **94**, 326–332 (2007)

Ants have opposable spines on their forelegs that allow them to manipulate objects, say researchers.

It was thought that ants used their legs for walking, and their jaws for carrying, nest-building, and so on. But Deby Cassill of the University of South Florida in St Petersburg and her colleagues saw queens of the fire ant *Solenopsis invicta* (pictured below) manipulating eggs and larvae with their forelegs. The team used electron microscopy to reveal the pincer-like paired spines on the limbs of several castes.

Ants join a short list of animals that use their forelimbs in this way, the team says, including crabs, koalas and primates. Seven of eight other ant species investigated had similar spines, suggesting that this might be a widespread feature.



PHOTOLIBRARY.COM

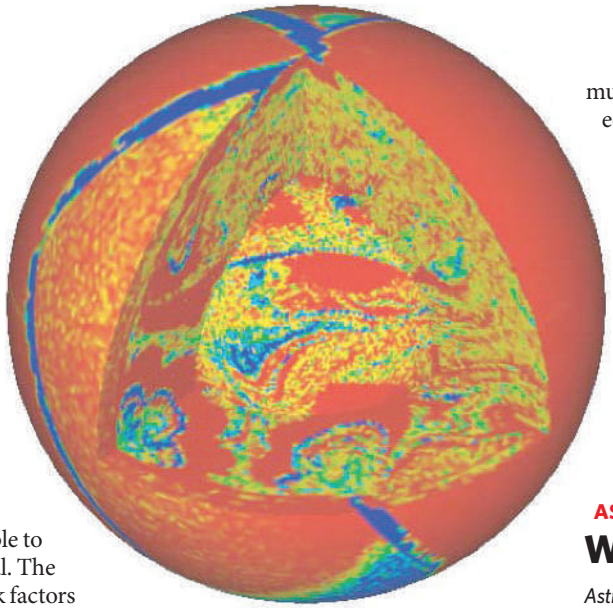
CHEMICAL BIOLOGY

Sensitive to shape

Angew. Chem. Int. Edn doi:10.1002/anie.200604995 (2007)

A simple chemical system can perform 'shape recognition', report Rustem Ismagilov of the University of Chicago, Illinois, and his colleagues.

Experiments with human blood plasma showed that blood clotting occurs when a chemical trigger, known as tissue factor, is deposited in patches with particular shape, but not when the same amount of tissue factor is laid down in other shapes. For example, clotting above rectangular patches only occurred when the ratio of the lengths of the rectangle's sides met a particular criterion.



G. DAVIES, AUSTRALIAN NATL UNIV.

The researchers reproduced this behaviour in a numerical simulation. They also demonstrated shape sensitivity in a chemical system that mimics the properties of the blood-clotting network. Further work is planned to determine the effect's biological relevance.

GENETICS

Change of heart

J. Am. Med. Assoc. **297,** 1551–1561 (2007)

A study looking at 85 genetic variations previously linked to heart disease — some of which are already used in clinical tests — has been unable to confirm that any of these links are real. The findings highlight the need to test risk factors in large populations, says the team that did the work.

Thomas Morgan of the Washington University School of Medicine in St Louis, Missouri, and his colleagues tested 1,461 Caucasian patients for genetic risk factors reported in the literature. They found no significant association in their patients between any of the specific gene variants and development of acute coronary syndrome. This does not rule out the variants as risk factors for some groups, but it emphasizes the need for caution in applying the findings of genetic studies.

GEOPHYSICS

Stirring things up

Geochem. Geophys. Geosyst. **8,** Q03017 (2007)

A three-dimensional model of melting and mixing in Earth's mantle confirms that all but about 3% of this rocky layer has been melted at some point in the past.

Mantle melts as it wells up at mid-ocean ridges to form new crust, which eventually sinks back into the mantle at deep ocean trenches and gets mixed back into the whole through a process known as convection. Understanding this cycle is important for understanding how Earth's interior and atmosphere evolved.

The computer model (pictured above) created by Jinshui Huang and Geoffrey Davies at the Australian National University in Canberra also shows how chemical patchiness in the mantle survives for about 2 billion years in spite of mixing.

STEM CELLS

Gender matters

J. Cell Biol. **177,** 73–86 (2007)

The differences between men and women may extend all the way down to their stem cells, report Johnny Huard of the University of Pittsburgh, Pennsylvania, and his colleagues.

The researchers found that female muscle-derived stem cells are better at regenerating muscle in a mouse model than their male equivalents. They saw that male stem cells differentiate into muscle more quickly when put under stress than female cells, which may limit the total number of muscle fibres that form because it leaves the stem cells little time to multiply. But the underlying mechanism remains unclear.

The team urges other researchers to report the sex of stem cells used in experiments, in case it turns out to influence the behaviour of other types of stem cell too.

ASTRONOMY

Water seen by starlight

Astrophys. J. (in the press) preprint at http://lanl.arxiv.org/abs/0704.1114 (2007)

An astronomer in the United States claims to have detected water in the atmosphere of a planet far, far away.

Travis Barman at Lowell Observatory in Flagstaff, Arizona, compared theoretical models of the atmosphere of the extrasolar planet HD209458b — a hot gas giant some 150 light years from Earth — with observations from the Hubble Space Telescope. The models best fit the data if water is present, he says, providing the first hints of water on a planet beyond our Solar System.

Earlier this year, researchers using a different technique to study the same planet reported no evidence of water (L. J. Richardson *et al. Nature* **445,** 892–895; 2007). But both teams say the results are not necessarily contradictory; Barman's models focused on the effect of the planet's atmosphere on starlight as the planet passed in front of its parent star, whereas the previous work looked at radiation emitted by the planet during its daytime, when water could be harder to detect.

JOURNAL CLUB

Pablo Debenedetti
Princeton University, New Jersey, USA

A chemical engineer is struck by the strange properties of 'patchy' colloids.

A recent paper about the behaviour of colloids makes an intriguing prediction — suggesting that they can adopt an 'empty' liquid state.

I study disordered states of matter, such as liquids and glasses. I find colloids interesting because they make phenomena such as crystal nucleation and the glass transition amenable to direct observation. Nanometre- or micrometre-sized particles suspended in liquids are wonderful model atoms. They arrange themselves in the same way that atoms and simple molecules do into solids, liquids or gases.

But controlling the interactions between colloidal particles provides a window into structural and thermodynamic behaviour beyond that found in atomic systems, as this recent theoretical paper shows (E. Bianchi *et al. Phys. Rev. Lett.* **97,** 168301; 2006).

It maps the phase diagrams of 'patchy' colloids. The particles in such colloids are decorated with sticky spots, which tend to bond them together. As the number of bonded neighbours per particle is reduced towards two, the phase diagrams predict liquid states with a vanishing packing fraction. This means the colloidal particles occupy a tiny fraction of the available space — but they still behave as a liquid that is distinct from the gas-like phase of still lower packing fraction.

The low-temperature behaviour of such 'empty' liquids is especially interesting. The calculations suggest that cooling the colloid can freeze in place the empty configuration to give a glassy state of arbitrarily low density.

These predictions have not been tested experimentally. But chemists have already developed techniques for making patchy particles, so the work of Bianchi *et al.* could guide experimentalists in their exploration of this fascinating form of matter.

NEWS

Colliders race for the Higgs

JACKSONVILLE, FLORIDA

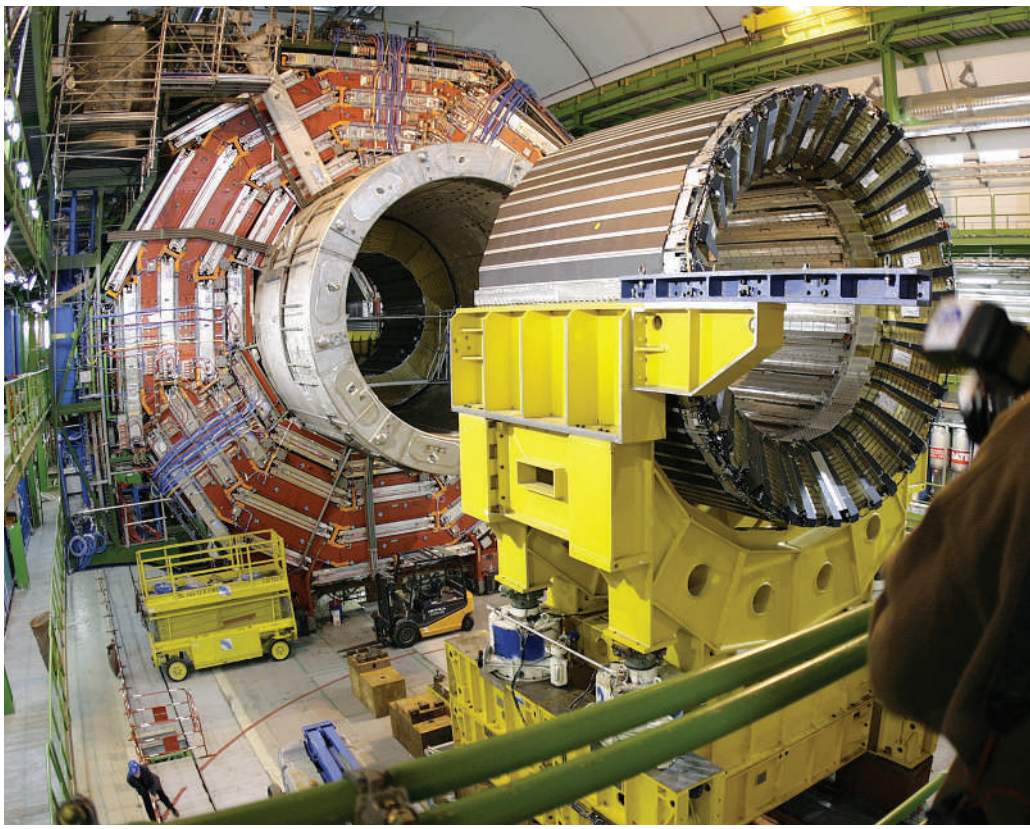
Soon after the news emerged on 29 March that a magnet test in the Large Hadron Collider (LHC) atom-smasher had failed, two physics bloggers decided to have a bit of fun with the idea. On 1 April they posted blogs announcing news such as 'Three years of delay for LHC start-up'. Visitors found the reports all too easy to swallow, decrying it as a cruel blow. Only then did one blogger add an April Fool's spoiler, with an apology: "Sorry to those of you who got hurt by not understanding that in the first place."

If these blog postings were close to the bone for many, it is because the LHC is running hard up against its deadline to switch on later this year. Conspiracy theories were also quick to fly: any postponement at the LHC, near Geneva, Switzerland, means that the Tevatron collider at the Fermi National Accelerator Laboratory in Illinois has more time to discover the last part of the standard model of particle physics — the Higgs boson.

On 15 April, the race became even more heated. At a meeting of the American Physical Society in Jacksonville, Florida, Tevatron physicists announced refined measurements for the masses of two particles that, taken together, lower the expected mass of the Higgs. The new estimates — of the mass of the top quark and of the W boson — put an upper limit on the Higgs mass of 144 GeV. In January, the best estimate was 153 GeV (see *Nature* **445,** 239; 2007). The lighter the Higgs, the better the chance the Tevatron has to detect it before it shuts down in 2009.

Researchers are more immediately worried about potential LHC delays caused by the failure of the magnet, which was supplied by Fermilab. "This does add to the burden of everything that has to be done before the machine switches on," says Pier Oddone, director of Fermilab. "We are embarrassed we created this additional problem." At the same time, some LHC researchers admit privately that the mistake might give them some breathing room.

The magnet that failed was part of an inner triplet designed to focus the proton beams before they interact. During a pressure test to simulate conditions expected in the collider, the eight-tonne magnet jumped 13 centimetres, rupturing a pipe and causing a loud bang. The problem was quickly identified as a weakness in the supporting structure. "There was a definite oversight here," admits Stephen Holmes, head of accelerator physics at Fermilab.



Big bang: the experiments at Europe's Large Hadron Collider are unlikely to be running before 2008.

AP PHOTO/M. TREZZIN

The accident was not the first magnet failure: another Fermilab-supplied magnet had a faulty heat exchanger that had to be replaced a few months earlier. The second problem was one too many for Oddone, who has initiated an external review to figure out how the team missed such simple design flaws.

Meanwhile, CERN — the particle-physics laboratory in which the LHC is housed — and Fermilab are working together to find a fix and to minimize the effect on the LHC schedule. Options include reinforcing the triplet support structure with rods, or building physical buffers so that the magnets cannot move too far. Engineers hope to announce a solution by 25 April, and to test it on another triplet in the tunnel in early June.

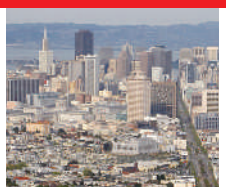
> **"The magnet problem is a very small part of a bigger picture."**

The time needed to repair the magnets will need to be factored into the overall LHC operation, which was already running five weeks behind schedule before the accident, says project manager Lyndon Evans. There is also pressure to get the repair right the first time around, because there are few spare magnets available.

The most likely victim of any slippage is the first engineering run, planned for late 2007. The LHC is being chilled down in sections; it took from January to March 2007 to cool the first eighth of the machine to 1.9 kelvin. The process will become quicker, but cooling the final sectors in just two weeks each — as Evans says the current schedule demands — will be challenging, with or without the magnet repairs. "In my view the magnet problem has been blown out of proportion," he says. "It is a very small part of a bigger picture."

Evans adds that he still hasn't given up on the LHC conducting its first run in 2007, but admits that only an extreme optimist would share his view. With delay building on delay, a formal announcement of when the LHC might come online isn't expected until mid-May. And with CERN closing as it usually does for the winter months, no run in 2007 could push the first science run later than planned in 2008. And by that point, the Tevatron will have had another year or more to hunt for the Higgs. ■

Sarah Tomlin



BIG CITIES NEED A FAST-PACED LIFE TO GROW
Huge urban centres are fed by rapid innovation.
www.nature.com/news

Interim view from NASA relativity probe

Gravity Probe B (GP-B) seems like the experiment that never ends. The spacecraft set records for the longest-running development project at NASA — 40 years — and data analysis is stretching into its second year. The GP-B team may one day announce the most precise measurements yet of a long-sought effect of general relativity. But that didn't happen on 14 April, when project scientists presented an interim report to the American Physical Society meeting in Jacksonville, Florida.

Team members promise a final report by December, when money for the $760-million experiment runs out. But it is clear that unexpected systematic errors will make it a real challenge to reach the original mission goals. For the most subtle effect measured, the GP-B team needs to make the experiment's uncertainty 100 times lower.

GP-B is a simple concept that in reality proved overwhelming. The experiment, proposed in 1964, required four perfect spinning gyroscopes in Earth orbit to measure how the spinning planet drags the fabric of space-time around with it — a phenomenon predicted by Einstein's general theory of relativity and called 'frame-dragging'. The mission survived every NASA attempt to cancel it (see *Nature* 426, 380–381; 2003) and was finally launched in April 2004.

K. STEPHENSON, STANFORD UNIV., LOCKHEED MARTIN CORP.

NASA's Gravity Probe B aims to confirm two predictions of general relativity.

The main goal was to measure frame-dragging to within 1% accuracy. That would be ten times better than the best measurements so far, which were taken by bouncing laser signals off the Earth-orbiting LAGEOS satellites.

GP-B has had one success. At the meeting, the team announced the first experimental measurement of geodetic precession, another small distortion of space-time. The geodetic precession of the Moon around the Earth has previously been verified to an accuracy of about 0.7%. GP-B has now measured the effect on their probe to within 1.5%. The team hopes to reduce this further — but it is the much smaller frame-dragging effect that everyone cares about.

The problems plaguing the analysis are systematic errors: electrostatic effects on the spheres at the core of the gyroscopes cause misalignments and wobbles that vary unexpectedly over time. The data were also recorded in chunks because solar flares required the system to be rebooted.

Physicist Clifford Will, who chairs NASA's scientific advisory committee for the project, says such errors are a real headache: "There's art involved — it's a slightly nebulous business; that's why they are being careful." Still, the team remains bullish about the remaining data analysis. "I'm not interested in being disappointed," says Francis Everitt of Stanford University, the project's principal investigator.

Meanwhile, the LAGEOS findings could be further improved before the end of the year by incorporating a new model of Earth's gravitational field, as gathered by the GRACE spacecraft. Erricos Pavlis of the University of Maryland in Baltimore County, who works with the LAGEOS data, says that it would be nice to beat GP-B, but even so he doesn't want to see it fail.

"After all these millions spent and decades of people's work," he says, "it's only fair that they get something out of this project." ■

Sarah Tomlin

Evidence for fourth neutrino fades

A nine-year effort to resolve a mystery about the behaviour of neutrinos — particles that interact only weakly with matter — has thrown up an unexpected signal. The findings, announced on 11 April, leave open the possibility that new physics hides within the observations, but winnows the options for what that might be.

"The plot keeps thickening," says Bill Louis, co-spokesperson for the MiniBooNE neutrino detector based at Fermilab in Batavia, Illinois.

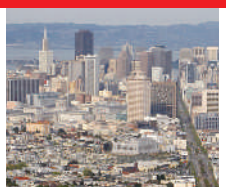
MiniBooNE aimed to settle controversy raised by the Liquid Scintillator Neutrino Detector (LSND) at Los Alamos National Laboratory in New Mexico, which looked at how one type of neutrino could turn into another, or 'oscillate'. The LSND findings suggested that a fourth kind of neutrino existed — a sterile neutrino. But many scientists were sceptical of the result.

Using a 12-metre sphere filled with 800 tonnes of mineral oil to catch neutrinos, the MiniBooNE team found no evidence, at high neutrino energies, for the sort of oscillation that workers at LSND had reported (A. A. Aguilar-Arevalo *et al.* preprint at http://arxiv.org/abs/0704.1500; 2007). "My view is that there is no longer any credible evidence for sterile neutrinos," says Gary Feldman, a physicist at Harvard University. Team members say they are still working on the data and can't be so sure.

At low neutrino energies the detector saw more electron neutrinos in the experiment's beam of muon neutrinos than expected. The team can't yet explain the observation and the community's curiosity is piqued. "Already theorists are sending us their papers and saying look, we fit you," says project co-spokesperson Janet Conrad.

Meanwhile, MiniBooNE has switched to observing antineutrinos, which is what LSND studied, to find out whether the variations between the two experiments' results are due to surprising differences in the behaviour of matter and antimatter. ■

Jenny Hogan



BIG CITIES NEED A FAST-PACED LIFE TO GROW
Huge urban centres are fed by rapid innovation.
www.nature.com/news

Interim view from NASA relativity probe

Gravity Probe B (GP-B) seems like the experiment that never ends. The spacecraft set records for the longest-running development project at NASA — 40 years — and data analysis is stretching into its second year. The GP-B team may one day announce the most precise measurements yet of a long-sought effect of general relativity. But that didn't happen on 14 April, when project scientists presented an interim report to the American Physical Society meeting in Jacksonville, Florida.

Team members promise a final report by December, when money for the $760-million experiment runs out. But it is clear that unexpected systematic errors will make it a real challenge to reach the original mission goals. For the most subtle effect measured, the GP-B team needs to make the experiment's uncertainty 100 times lower.

GP-B is a simple concept that in reality proved overwhelming. The experiment, proposed in 1964, required four perfect spinning gyroscopes in Earth orbit to measure how the spinning planet drags the fabric of space-time around with it — a phenomenon predicted by Einstein's general theory of relativity and called 'frame-dragging'. The mission survived every NASA attempt to cancel it (see *Nature* 426, 380–381; 2003) and was finally launched in April 2004.

K. STEPHENSON, STANFORD UNIV., LOCKHEED MARTIN CORP.

NASA's Gravity Probe B aims to confirm two predictions of general relativity.

The main goal was to measure frame-dragging to within 1% accuracy. That would be ten times better than the best measurements so far, which were taken by bouncing laser signals off the Earth-orbiting LAGEOS satellites.

GP-B has had one success. At the meeting, the team announced the first experimental measurement of geodetic precession, another small distortion of space-time. The geodetic precession of the Moon around the Earth has previously been verified to an accuracy of about 0.7%. GP-B has now measured the effect on their probe to within 1.5%. The team hopes to reduce this further — but it is the much smaller frame-dragging effect that everyone cares about.

The problems plaguing the analysis are systematic errors: electrostatic effects on the spheres at the core of the gyroscopes cause misalignments and wobbles that vary unexpectedly over time. The data were also recorded in chunks because solar flares required the system to be rebooted.

Physicist Clifford Will, who chairs NASA's scientific advisory committee for the project, says such errors are a real headache: "There's art involved — it's a slightly nebulous business; that's why they are being careful." Still, the team remains bullish about the remaining data analysis. "I'm not interested in being disappointed," says Francis Everitt of Stanford University, the project's principal investigator.

Meanwhile, the LAGEOS findings could be further improved before the end of the year by incorporating a new model of Earth's gravitational field, as gathered by the GRACE spacecraft. Erricos Pavlis of the University of Maryland in Baltimore County, who works with the LAGEOS data, says that it would be nice to beat GP-B, but even so he doesn't want to see it fail.

"After all these millions spent and decades of people's work," he says, "it's only fair that they get something out of this project." ■

Sarah Tomlin

Evidence for fourth neutrino fades

A nine-year effort to resolve a mystery about the behaviour of neutrinos — particles that interact only weakly with matter — has thrown up an unexpected signal. The findings, announced on 11 April, leave open the possibility that new physics hides within the observations, but winnows the options for what that might be.

"The plot keeps thickening," says Bill Louis, co-spokesperson for the MiniBooNE neutrino detector based at Fermilab in Batavia, Illinois.

MiniBooNE aimed to settle controversy raised by the Liquid Scintillator Neutrino Detector (LSND) at Los Alamos National Laboratory in New Mexico, which looked at how one type of neutrino could turn into another, or 'oscillate'. The LSND findings suggested that a fourth kind of neutrino existed — a sterile neutrino. But many scientists were sceptical of the result.

Using a 12-metre sphere filled with 800 tonnes of mineral oil to catch neutrinos, the MiniBooNE team found no evidence, at high neutrino energies, for the sort of oscillation that workers at LSND had reported (A. A. Aguilar-Arevalo *et al.* preprint at http://arxiv.org/abs/0704.1500; 2007). "My view is that there is no longer any credible evidence for sterile neutrinos," says Gary Feldman, a physicist at Harvard University. Team members say they are still working on the data and can't be so sure.

At low neutrino energies the detector saw more electron neutrinos in the experiment's beam of muon neutrinos than expected. The team can't yet explain the observation and the community's curiosity is piqued. "Already theorists are sending us their papers and saying look, we fit you," says project co-spokesperson Janet Conrad.

Meanwhile, MiniBooNE has switched to observing antineutrinos, which is what LSND studied, to find out whether the variations between the two experiments' results are due to surprising differences in the behaviour of matter and antimatter. ■

Jenny Hogan

Court case to reclaim confidential data

A public-health expert in the United States is facing a potential jail term or a hefty fine after distributing documents that contain safety data on a blockbuster drug. The material came to light during initial negotiations in an ongoing lawsuit over the drug, and its manufacturer says that the information should have remained confidential.

The case, which could result in charges of contempt being levelled against the expert, shines light on the murky debates that go on between lawyers before drug liability cases get to court. Legal experts say that the negotiations often end with crucial data being sealed from public view after lawyers have finished with it, even though they could be relevant to public health.

"Defendants essentially pay for the right to keep stuff secret," says Paul Levy, an attorney with Public Citizen, a consumer-advocacy organization based in Washington DC. "It is a conflict between public and private interests."

The latest incident centres around David Egilman from Brown University in Providence, Rhode Island, who is a frequent critic of the pharmaceutical industry, and Eli Lilly, a drug firm based in Indianapolis, Indiana. Lilly's best-earning drug is Zyprexa, a treatment for schizophrenia that brings in more than US$4 billion a year. But over the past two years Lilly has paid $1.2 billion to more than 26,000 patients who complained that the drug caused them to develop diabetes and hyperglycaemia. Other lawsuits related to Zyprexa and diabetes are ongoing, including one that Egilman last year agreed to contribute to as an expert witness.

Egilman had access to 15 million pages of confidential documents containing clinical-trial data and details of marketing plans that Lilly had released to lawyers during pre-trial negotiations, known as the discovery phase. But because the cases were settled before they reached court, the material was bound by confidentiality agreements.

This situation is common and has angered public-health experts and some lawyers. Discovery documents often contain evidence that suggests a drug is not as effective or as safe as its manufacturer claims and the information has sometimes not been released to regulators such as the US Food and Drug Administration. Yet lawyers are under no obligation to ask for the confidentiality order to be removed, even if the material suggests that the public is at risk.

"Defendants essentially pay for the right to keep stuff secret."

In the Zyprexa case, 376 documents emerged after what might be called creative journalism by the *New York Times*. According to a 13 February statement from Jack Weinstein, the New York district judge overseeing the case, Egilman discussed the documents with Alex Berenson, a *Times* reporter. To get the material, Berenson gave Egilman the name of an attorney who, the judge said, "would be a willing ally in an attempt to avoid the court's protective order by finding a case which could be used as a pretence for subpoenaing the protected documents". The attorney, James Gottstein of Anchorage, Alaska, subpoenaed Egilman for the Zyprexa documents and then forwarded them on to the reporter.

The result was a series of front-page stories in which Berenson alleged that Lilly had downplayed the risks that Zyprexa users would develop diabetes. The stories also claimed that Lilly tried to persuade doctors to prescribe the drug for dementia, a condition that it was not licensed for. Lilly spokesman Philip Belt declined to comment on either issue, saying only that the information quoted by Berenson was taken out of context.

Although Egilman was subpoenaed to release the documents, he and Gottstein could face jail terms or be required to contribute to Lilly's legal bills if the contempt proceedings proceed and are upheld. The documents, meanwhile, remain available online on a blog about mental-health issues.

Generally though, information uncovered during the discovery phase remains secret. In a separate instance, psychiatrist David Healy, of Cardiff University in Wales, says that he received confidential documents for several cases on which he was to act as an expert witness. Lawyers in the cases were acting for individuals who claimed that they or their relatives had had suicidal or homicidal thoughts after taking antidepressants known as selective seratonin reuptake inhibitors (SSRIs). All the documents had to go back to the firms that supplied them, but "there is a great deal of material that ought to be out there", says Healy.

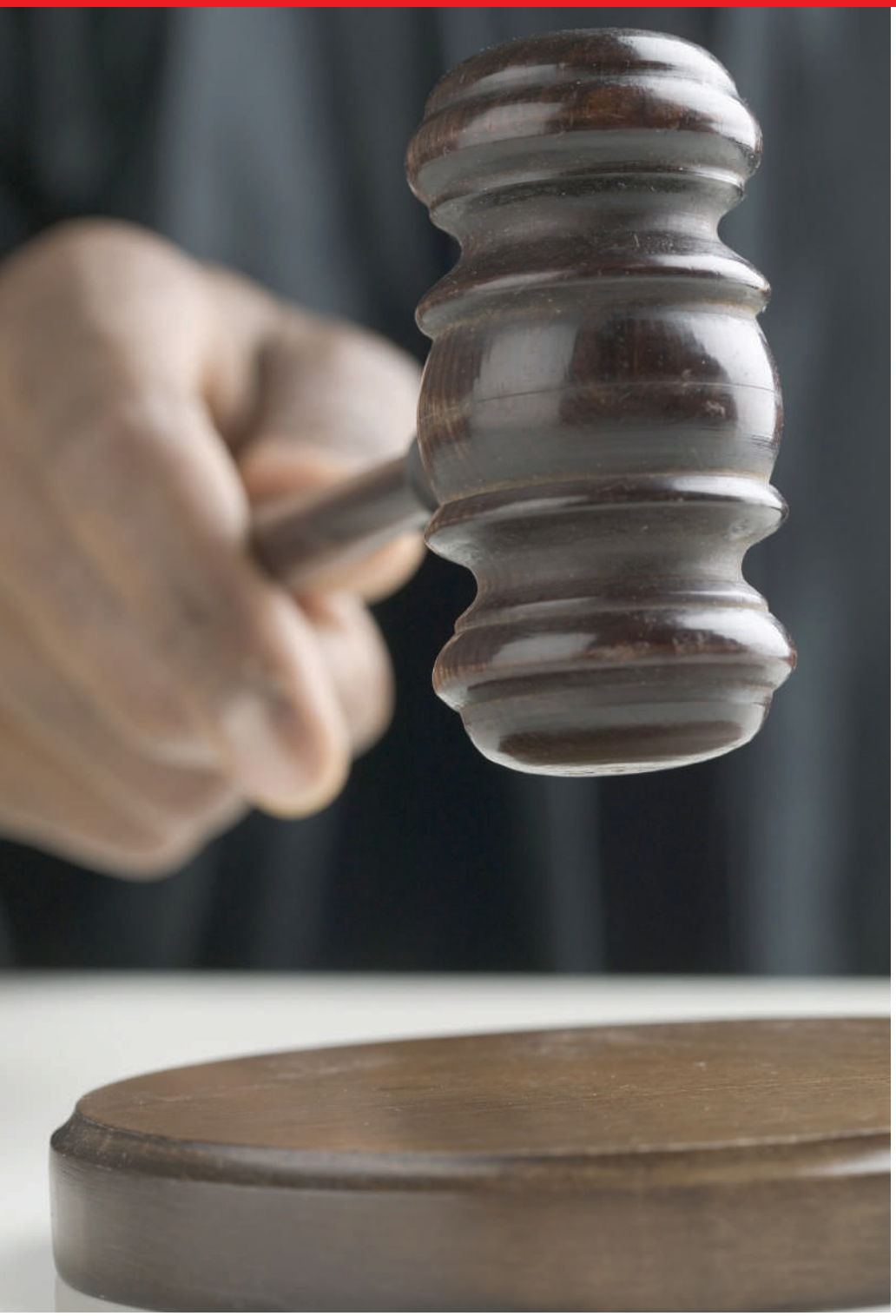
Documents have also been sealed in other SSRI cases that were settled before trial and in litigation involving prescription painkillers such as oxycontin and neurontin. And in 2004, publication of a study on cancer risk at an IBM factory was delayed after lawyers for the firm argued that the data were covered by a confidentiality order. "Concealment of such evidence corrupts the scientific literature," says Vera Sharav of the Alliance for Human Research Protection, a patient advocacy group based in New York. (Sharav had also received the documents from Gottstein, but was ordered by Weinstein to return them).

Changing the rules governing discovery documents will be difficult. Discovery material often contains trade secrets, such as details of how a drug is made, as well as records for individual patients. Without confidentiality, defendants would be much more reluctant to give up such materials. "The judge knows that the secrecy is the grease that keeps things moving," says Gottstein. "He doesn't want the settlement to get gummed up."

Critics acknowledge that some material needs to be kept under wraps, but that this should not prevent documents from being released if they are in the public interest. Some legal firms, such as Baum Hedlund in Los Angeles, California,

D. CUMMINGS/AP

The schizophrenia drug Zyprexa has rekindled a debate about the confidentiality of documents involved in lawsuits.



CORBIS

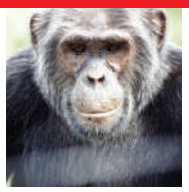
Information revealed during court trials can be sealed from public view even if in the public interest.

devote time to challenging protection orders alongside pursuing their obligations to their clients. Others add that plaintiffs sometimes want the material to be made openly available in the interest of the public, and will ask the lawyers not to sign confidentiality agreements. Documents can also come out if politicians ask for them: Senator Chuck Grassley (Republican, Iowa) wrote to Lilly on 4 April asking to see all the Zyprexa material.

But the chances for broader reform seem to be limited. Richard Zitrin, director of the Center for Legal Ethics at the University of San Francisco in California, says that legislation to prevent lawyers from signing confidentiality agreements that go against public safety has failed several times because of pressure from corporate lobby groups, including those representing the pharmaceutical industry.

In the meantime, attorneys will continue to have to seal documents that they know should be made public, something many find intensely frustrating. "It burns you up inside," says Cindy Hall, a pharmaceutical litigation consultant with Baum Hedlund. ■

Jim Giles



WONDERLUSTG

CHIMP GENOME
Learn about primate genomes at our chimp genome special
www.nature.com/news/specials/chimpgenome

Ozone sensor to be reinstated

Fans of ozone data, rejoice. US science agencies announced on 11 April that they will restore an axed ozone sensor to a satellite that is due to launch in 2009.

The sensor was scrapped last June, when the National Polar-orbiting Operational Environmental Satellite System (NPOESS) was brutally restructured. The system is meant to merge and replace the US military and civilian weather-prediction satellites, but it is over budget and behind schedule. At least three climate sensors were axed at the time.

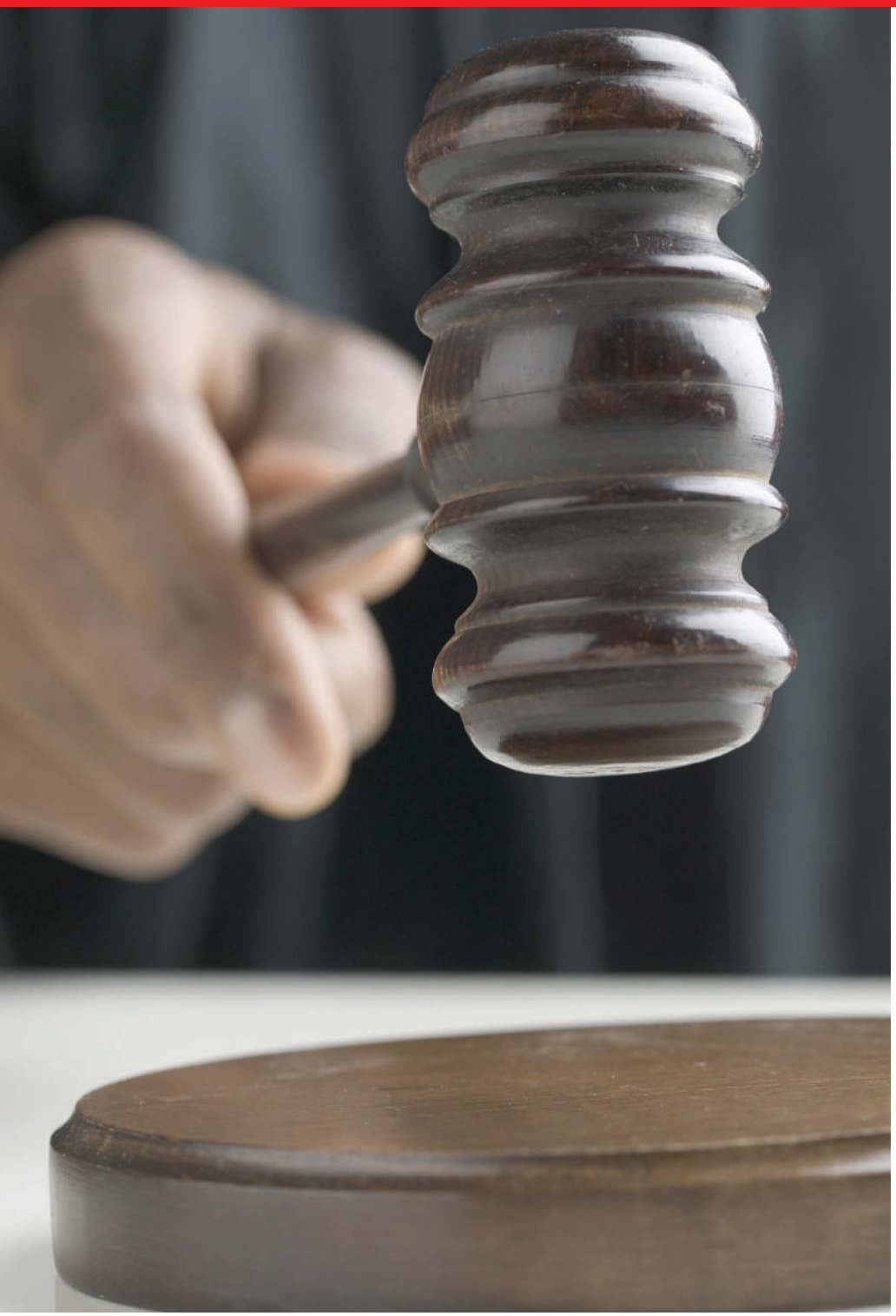
Now, the instrument known as the Ozone Mapping and Profiler Suite Limb will get to fly aboard the initial satellite in the system, according to NASA and the National Oceanic and Atmospheric Administration (NOAA). The sensor will measure the distribution of ozone concentrations from the ground to the top of the atmosphere, and can be used to monitor the ozone layer and the slow recovery of the ozone hole.

Earth scientists at NASA, along with ozone scientists, pushed hard for the sensor to be reinstated. Pawan Bhartia, senior staff scientist at the Goddard Space Flight Center in Greenbelt, Maryland, who helped to develop the instrument, says that the sensor will be there to collect ozone data when the current satellites Aura and Envisat stop operating, perhaps in the next couple of years. "Right now, the missions we have cost a billion dollars each," he says. "But once you understand the science, you can keep on monitoring it on a long-term basis, but not spend so much money on it."

Andrew Carson, an executive for the project at NASA headquarters in Washington DC, says that the change of heart was influenced in part by a January 2007 National Academy of Sciences report that called for stronger Earth-monitoring efforts from space. But he says there were moves to restore the sensor even before the report came out.

The money will come from outside the main NPOESS budget, and will be split equally between NOAA and NASA. None of the other sensors cut last year are slated to be restored. ■

Emma Marris



CORBIS

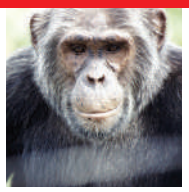
Information revealed during court trials can be sealed from public view even if in the public interest.

devote time to challenging protection orders alongside pursuing their obligations to their clients. Others add that plaintiffs sometimes want the material to be made openly available in the interest of the public, and will ask the lawyers not to sign confidentiality agreements. Documents can also come out if politicians ask for them: Senator Chuck Grassley (Republican, Iowa) wrote to Lilly on 4 April asking to see all the Zyprexa material.

But the chances for broader reform seem to be limited. Richard Zitrin, director of the Center for Legal Ethics at the University of San Francisco in California, says that legislation to prevent lawyers from signing confidentiality agreements that go against public safety has failed several times because of pressure from corporate lobby groups, including those representing the pharmaceutical industry.

In the meantime, attorneys will continue to have to seal documents that they know should be made public, something many find intensely frustrating. "It burns you up inside," says Cindy Hall, a pharmaceutical litigation consultant with Baum Hedlund. ■

Jim Giles



WONDERLUSTG

CHIMP GENOME
Learn about primate genomes at our chimp genome special
www.nature.com/news/specials/chimpgenome

Ozone sensor to be reinstated

Fans of ozone data, rejoice. US science agencies announced on 11 April that they will restore an axed ozone sensor to a satellite that is due to launch in 2009.

The sensor was scrapped last June, when the National Polar-orbiting Operational Environmental Satellite System (NPOESS) was brutally restructured. The system is meant to merge and replace the US military and civilian weather-prediction satellites, but it is over budget and behind schedule. At least three climate sensors were axed at the time.

Now, the instrument known as the Ozone Mapping and Profiler Suite Limb will get to fly aboard the initial satellite in the system, according to NASA and the National Oceanic and Atmospheric Administration (NOAA). The sensor will measure the distribution of ozone concentrations from the ground to the top of the atmosphere, and can be used to monitor the ozone layer and the slow recovery of the ozone hole.

Earth scientists at NASA, along with ozone scientists, pushed hard for the sensor to be reinstated. Pawan Bhartia, senior staff scientist at the Goddard Space Flight Center in Greenbelt, Maryland, who helped to develop the instrument, says that the sensor will be there to collect ozone data when the current satellites Aura and Envisat stop operating, perhaps in the next couple of years. "Right now, the missions we have cost a billion dollars each," he says. "But once you understand the science, you can keep on monitoring it on a long-term basis, but not spend so much money on it."

Andrew Carson, an executive for the project at NASA headquarters in Washington DC, says that the change of heart was influenced in part by a January 2007 National Academy of Sciences report that called for stronger Earth-monitoring efforts from space. But he says there were moves to restore the sensor even before the report came out.

The money will come from outside the main NPOESS budget, and will be split equally between NOAA and NASA. None of the other sensors cut last year are slated to be restored. ■

Emma Marris

Make way for monkeys

Rhesus macaques can't claim the distinction of being humanity's closest cousins, nor of ever having been popularized by Jane Goodall. But the newly completed analysis of the macaque's genome could prove just as important as that of the chimpanzee's, scientists say.

Macaca mulatta is the third primate genome to be sequenced, following those for humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). Adding the macaque to that picture, scientists say, should allow researchers to make better sense of the other two genomes. And because the rhesus macaque is the main monkey model for many diseases, understanding its genome could help to clarify why some diseases develop, and why some treatments work differently in monkeys from in humans.

The macaque genome, published on 13 April in *Science*, is crucial for unravelling the evolutionary events that led to humanity. Ever since scientists released the chimp genome sequence in 2005 (see *Nature* **437,** 69–87; 2005), they have been busily trying to dissect how it differs from our own — looking for clues to explain uniquely human traits.

But when a genetic difference is identified between humans and chimps, it is not always possible to tell which form of the DNA is older. Scientists can compare each sequence to those of distantly related animals, such as chickens and mice, but this technique fails when the genetic variant is so new that it is unique to primates. The macaque is therefore an ideal reference point, scientists say, because the evolutionary split between monkeys and apes (such as humans and chimps) occurred 25 million years ago, much earlier than the human–chimp divergence 6 million years ago.

"Now that we have a third primate, a large part of the focus is understanding the changes that happened in each primate lineage, and that is very important in understanding evolution," says George Weinstock of the Baylor College of Medicine in Houston, Texas, a leader of the macaque sequencing effort. The project involved researchers from North America, Asia and Europe (Rhesus Macaque Genome Sequencing and Analysis Consortium *Science* **316,** 222–234; 2007).

"Understanding the changes in each primate lineage is very important in understanding evolution."

And the macaques could soon have company. Other primate genome sequences in progress include the gibbon (*Nomascus leucogenys*), orangutan (*Pongo pygmaeus*), gorilla (*Gorilla gorilla*) and marmoset (*Callithrix jacchus*).

B. CASTELEIN/NATURE PICTURE LIBRARY

Analysis of the macaque genome could lead to improved models for diseases.

The new analysis finds that where humans and macaques share genes, the DNA sequences in those genes are 97.5% identical. But the findings also highlight differences between the species' DNA. For instance, the scientists report that the 108 gene families shared between humans, macaques and chimps have evolved differently in the three species, and about 60% of those amplified in the macaque show signs of natural selection. Some of these genes probably explain differences between humans and monkeys — like our inability to swing through forest treetops and our revulsion to ingesting our own vomit.

The differences might also shed light on more medical issues. Macaques are often used as stand-ins for humans in experimental tests of drugs and medical treatments — especially for infectious diseases, such as HIV. But these monkey tests do not always accurately predict how the treatments behave in the human body.

The new analysis finds that macaques have many more copies of genes corresponding to those making up the human HLA system — a key part of the body's defence against diseases. Finding that the HLA system is genetically different in macaques could be a sign that the monkeys' immune systems don't work like our own. That information, in turn, could help researchers design more accurate preclinical tests in macaques.

"A lot of people use macaques as a model for the immune response, so the fact there are these differences is important," says Evan Eichler of the University of Washington in Seattle, one of the leaders of the analysis portion of the macaque project.

And that, scientists say, means that the macaque genome should be analysed in even more detail. The current publication is only a 'draft' sequence — meaning it is not as complete as, for instance, the human genome.

Historical hangovers

But already, plenty of material is available to chew on. For instance, the analysis turned up hundreds of cases in which humans and macaques share a gene, but although the macaque version keeps the monkey healthy, the human one causes disease. Understanding the implications of this finding could take some time. But it may cause some scientists to rethink their approach to the nature of disease. Rather than being a recent glitch in our physiology, a disease-causing gene might actually be a holdover from our ancestral history that hasn't kept pace with the way our bodies or lifestyles have evolved, and has now become detrimental.

"We tend to think about disease very simplistically, that everything was fine until a new variant was introduced, and then things went awry," Weinstock says. "But actually, it is much more complex than that." ■

Erika Check

Chimps lead evolutionary race

Humans are generally believed to be more highly 'evolved' than our chimpanzee cousins. But in at least one sense that isn't true, say geneticists who have hunted for the hallmarks of natural selection in our respective genomes — and found more of them in chimps.

The discovery suggests that, since our evolutionary paths diverged 6 million years ago, greater numbers of chimpanzee genes have been shaped by 'positive selection', in which natural selection favours beneficial mutations.

Researchers at the University of Michigan, Ann Arbor, combed through 14,000 matching genes from the human and chimpanzee genomes. As they report in *Proceedings of the National Academy of Sciences* this week, 233 chimp genes showed signs of having been shaped by positive selection (M. A. Bakewell, P. Shi and J. Zhang *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0701705104; 2007). The corresponding figure for our own genes was just 154.

The result overturns the view that, to promote humans to our current position as the dominant animal on the planet, we must have encountered considerable positive selection, says lead author Jianzhi Zhang. "We think we're very different from animals, with our large brain size and speech," he says.

The gene discrepancy might be due to the fact that, for much of our histories, chimpanzees had the larger population size. Humans, with a smaller and more fragmented population, may have been shaped by random, erratic changes.

It is difficult to put together a coherent picture, says Zhang, because it is hard to know which genes would have been crucial in shaping traits such as our large brain size. "It is possible that the genetic changes underlying brain size are very few," he says.

A sample of 14,000 genes does not tell the whole story. The team could not compare the entire genome as the chimp sequence has not been completed to the same level of detail as the human one. But for genes with good sequences, they were taken to show signs of positive selection if they had a high proportion of 'non-synonymous mutations' — DNA changes that alter the protein sequence produced by the gene — which could be a 'lever' for natural selection.

"It is possible that the genetic changes underlying brain size are very few."

Zhang admits it is difficult to spot genes that have been the subject of more recent positive selection. Such genes could have been responding to selection pressures — such as changes in climate and food source — encountered by humans as they began to move out of Africa and across the planet over the past 100,000 years.

There also seems to be little pattern to the functions of the selected genes, says Zhang. Among those favoured in chimps are genes for protein metabolism and stress responses, whereas the human genes are involved in processes such as fatty-acid metabolism.

Victoria Horner, who works with chimpanzees at Yerkes National Primate Research Center in Atlanta, Georgia, says: "We assume that chimpanzees have changed less than us, when that's actually not the case." ■

Michael Hopkin

R. E. KLEIN/AP

Chimpanzees have at least 233 genes thought to be shaped by selection for beneficial mutations.

ZOO NEWS

Dial-up dolphin
Castaway, a deaf dolphin at the Marine Mammal Conservancy in Key Largo, Florida, has become the proud recipient of a 'chat line' that will broadcast the sounds of nearby dolphins and help the unborn calf in her uterus learn vital sonar and social skills.

SCORECARD

▲ **Durian fruit**
A Thai scientist says he has bred an odourless version of the pongy fruit, which is banned from many Asian hotels and airlines.

▼ **Pollen**
A dusting of pollen on cars and buildings makes it harder for police to find fingerprints left by thieves, crime experts have discovered.

ALAMY

NUMBER CRUNCH

8% is the amount that New York City's greenhouse-gas emissions have risen between 1995 and 2005.

58.3 million tonnes is the amount of greenhouse gas emitted by the city in 2005 — almost as much as the entire Republic of Ireland, but just 1% of total US emissions.

2.7% is the proportion of the US population that lives in the city — making New Yorkers almost three times greener, on average, than the typical American.

ON THE RECORD

"It would be the wrong thing to do to select people who aren't going to fit in your spaceship."

Duane Ross demonstrates exactly why he is NASA's head of astronaut selection.

Sources: ABC News, Associated Press, The Independent, USA Today

SIDELINES

Chimps lead evolutionary race

Humans are generally believed to be more highly 'evolved' than our chimpanzee cousins. But in at least one sense that isn't true, say geneticists who have hunted for the hallmarks of natural selection in our respective genomes — and found more of them in chimps.

The discovery suggests that, since our evolutionary paths diverged 6 million years ago, greater numbers of chimpanzee genes have been shaped by 'positive selection', in which natural selection favours beneficial mutations.

Researchers at the University of Michigan, Ann Arbor, combed through 14,000 matching genes from the human and chimpanzee genomes. As they report in *Proceedings of the National Academy of Sciences* this week, 233 chimp genes showed signs of having been shaped by positive selection (M. A. Bakewell, P. Shi and J. Zhang *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0701705104; 2007). The corresponding figure for our own genes was just 154.

The result overturns the view that, to promote humans to our current position as the dominant animal on the planet, we must have encountered considerable positive selection, says lead author Jianzhi Zhang. "We think we're very different from animals, with our large brain size and speech," he says.

The gene discrepancy might be due to the fact that, for much of our histories, chimpanzees had the larger population size. Humans, with a smaller and more fragmented population, may have been shaped by random, erratic changes.

It is difficult to put together a coherent picture, says Zhang, because it is hard to know which genes would have been crucial in shaping traits such as our large brain size. "It is possible that the genetic changes underlying brain size are very few," he says.

A sample of 14,000 genes does not tell the whole story. The team could not compare the entire genome as the chimp sequence has not been completed to the same level of detail as the human one. But for genes with good sequences, they were taken to show signs of positive selection if they had a high proportion of 'non-synonymous mutations' — DNA changes that alter the protein sequence produced by the gene — which could be a 'lever' for natural selection.

"It is possible that the genetic changes underlying brain size are very few."

Zhang admits it is difficult to spot genes that have been the subject of more recent positive selection. Such genes could have been responding to selection pressures — such as changes in climate and food source — encountered by humans as they began to move out of Africa and across the planet over the past 100,000 years.

There also seems to be little pattern to the functions of the selected genes, says Zhang. Among those favoured in chimps are genes for protein metabolism and stress responses, whereas the human genes are involved in processes such as fatty-acid metabolism.

Victoria Horner, who works with chimpanzees at Yerkes National Primate Research Center in Atlanta, Georgia, says: "We assume that chimpanzees have changed less than us, when that's actually not the case." ■

Michael Hopkin

R. E. KLEIN/AP

Chimpanzees have at least 233 genes thought to be shaped by selection for beneficial mutations.

ZOO NEWS

Dial-up dolphin
Castaway, a deaf dolphin at the Marine Mammal Conservancy in Key Largo, Florida, has become the proud recipient of a 'chat line' that will broadcast the sounds of nearby dolphins and help the unborn calf in her uterus learn vital sonar and social skills.

SCORECARD

▲ **Durian fruit**
A Thai scientist says he has bred an odourless version of the pongy fruit, which is banned from many Asian hotels and airlines.

▼ **Pollen**
A dusting of pollen on cars and buildings makes it harder for police to find fingerprints left by thieves, crime experts have discovered.

ALAMY

NUMBER CRUNCH

8% is the amount that New York City's greenhouse-gas emissions have risen between 1995 and 2005.

58.3 million tonnes is the amount of greenhouse gas emitted by the city in 2005 — almost as much as the entire Republic of Ireland, but just 1% of total US emissions.

2.7% is the proportion of the US population that lives in the city — making New Yorkers almost three times greener, on average, than the typical American.

ON THE RECORD

"It would be the wrong thing to do to select people who aren't going to fit in your spaceship."

Duane Ross demonstrates exactly why he is NASA's head of astronaut selection.

Sources: ABC News, Associated Press, The Independent, USA Today

SIDELINES

Stem-cell issue moves up the US agenda

WASHINGTON DC

US senators voted last week, for the second time in nine months, to lift restrictions on federal funding for human embryonic stem-cell research — even though they expect President George W. Bush to veto the decision again.

So why the repeat of last July's political meltdown? In part, senators are positioning themselves for next year's congressional elections and for the one political contest that will bear enormously on future US stem-cell policy: the presidential race of 2008.

In that race, stem-cell research "is going to be a very big issue, maybe right up alongside the war", says Daniel Greenberg, a veteran Washington science commmentator.

Jockeying began in the Senate on 11 April, when members voted 63–34 in favour of the Stem Cell Research Enhancement Act of 2007. The bill allows federal funding for research on stem cells that are derived from embryos left over at fertility clinics and already slated for destruction. It is similar to one passed by the Senate last summer and then vetoed by Bush (see *Nature* **442,** 335; 2006). And no sooner had last week's vote been cast than the president issued a statement promising to do the same again. "This bill crosses a moral line that I and many others find troubling," he said. "If it advances all the way through Congress to my desk, I will veto it."

The House of Representatives is expected to vote on the bill in the next few weeks. If it passes the House and is vetoed by Bush, it is unlikely that Congress will muster the two-thirds majority necessary in both houses to override the presidential veto. Three Democrats missed last week's Senate vote; all would almost certainly have voted in favour, bringing the total votes in favour to 66 — one shy of a two-thirds majority. The margin in the House is larger: a House vote on a similar bill in January failed by 32 votes to reach a veto-proof majority (see *Nature* **445,** 134–135; 2007).

So Bush's current policy, which dates from 9 August 2001, looks almost certain to remain in place for now. This limits federal funding to research on stem-cell lines that existed on that date, a number that has dwindled to 21.

But whoever assumes the presidency in January 2009 could implement his or her new stem-cell policy with "a wave of a hand" by using an executive order, says Jonathan Moreno, a professor of medical ethics at the University of Pennsylvania in Philadelphia. That means that voters on both sides of the issue are closely scrutinizing the positions of the leading presidential contenders.

A. C. GLENN/UPI/NEWSCOM

Tom Harkin (left), Arlen Specter and Orrin Hatch have pushed hard for changes to the stem-cell law.

"In the presidential race, stem-cell research will be a very big issue, maybe right up alongside the war."

Senator Sam Brownback (Republican, Kansas) spoke out against the bill last week and is the only declared presidential candidate who aggressively opposes stem-cell research. But he is a second-tier contender. Three other, stronger, Republican hopefuls are treading delicately as they try to court moderates without losing votes on the right.

Rudolph Giuliani, the Republican former mayor of New York City, has thus far skated around the stem-cell issue, saying that new technology needs to be taken advantage of while life also needs to be respected. Mitt Romney, a Mormon, originally supported stem-cell research as Republican governor of Massachusetts. Then in February 2005, he announced that he opposed the cloning of embryos for research purposes and would support criminal sanctions for those doing so. Senator John McCain (Republican, Arizona) voted in favour of the bill that passed last week.

So, too, did two leading Democratic contenders, Hillary Clinton of New York and Barack Obama of Illinois, both of whom were among the bill's many co-sponsors. And former senator John Edwards (Democrat, North Carolina) strongly supports federal funding for the research; after last week's Senate vote, he issued a statement urging Bush not to veto the bill.

Amongst all this political manoeuvring, Senate conservatives also bought themselves some political cover. Last week, the Senate passed the Hope Act, which Bush has promised to sign if it is also passed by the House. This allows federal funding for research on stem cells that are derived from embryos that the bill terms "naturally dead". The bill passed by a vote of 70 to 28, with Republicans the main supporters.

Sponsored by Norm Coleman (Republican, Minnesota), the bill defines "naturally dead" as "having naturally and irreversibly lost the capacity for integrated cellular division, growth, and differentiation". Critics, however, note that there is no scientific standard for determining what is a viable embryo. ■

Meredith Wadman

Cascade of errors ended space mission, inquiry finds

An official inquiry has blamed a computer-programming error for the loss of the Mars Global Surveyor spacecraft in November 2006.

In June 2006, new instructions sent to the spacecraft about what to do in certain special contingencies were mistakenly stored in the wrong memory address. Five months later, this led to a problem with the spacecraft's solar arrays. In responding to the problem, the spacecraft repositioned itself in such a way that one of its batteries overheated, while at the same time its main communication antenna turned away from Earth. The combination proved fatal.

The inquiry concludes that the spacecraft team had followed approved procedures, but that these weren't up to the task; equivalent procedures for other missions are being reviewed. The board also noted that cuts in budget and staff over the spacecraft's ten-year stint in orbit increased the risk that such things might happen.

Questions raised over 'cloned wolf' paper

The journal *Cloning and Stem Cells* has removed a paper announcing the cloning of wolves from its website (M. K. Kim *et al. Cloning and Stem Cells* 9, 130–137; 2007) pending the outcome of an investigation by Seoul National University.

The article, published in March, is the work of the same team at Seoul National University that was the first to clone dogs. The team had been led by Woo Suk Hwang, who is infamous for fraud in his reports of cloned human embryos, but its work to create the first cloned dog, Snuppy, was independently verified as authentic.

But is it a clone?

LEE JIN-MAN/AP

Following statements from the researchers that the paper needed corrections, a Seoul National University investigative committee was set up to test DNA samples from the cloned wolves, the wolf from which they were cloned, and the dogs that acted as surrogate mothers. Young Kuk, a physicist at the university who is leading the investigation, refused to comment on its progress. Byeong Chun Lee, who now leads the cloning group, is on trial for embezzlement charges related to the Hwang case.

US visa proves elusive for Iraqi researcher

An Iraqi researcher involved in a controversial study of post-invasion mortality rates has failed to get a visa to travel to the United States.

Riyadh Lafta of Al-Mustansiriya University in Baghdad was the sole Iraqi author on a paper published last October in *The Lancet* that claimed that between 390,000 and 940,000 people had died as a result of the invasion (see *Nature* 443, 728–729; 2006).

Last July, Lafta applied for a visa to travel to the University of Washington in Seattle, where researchers are planning a project on cancer and birth defects in Iraqi children. This March, the US Department of State said that it had offered Lafta a visa in October but had not received a reply and that the application had now expired. Lafta says he regularly checks the e-mail address the state department used and never received the message. He has now received a visa for Canada and will meet the Seattle researchers there.

Launch of physics journals boosts open-access club

Open-access publisher BioMed Central is launching three new physics journals under the sister brand-name PhysMath Central. They will sit alongside the company's portfolio of 176 biomedical titles.

The flagship title, *PMC Physics A*, will have as its editor-in-chief Ken Peach, director of the University of Oxford's John Adams Institute of Accelerator Science and chair of the scientific-policy committee at CERN, the European particle-physics laboratory near Geneva. It will focus on high-energy and nuclear physics, cosmology, gravity and astroparticle physics.

PMC Physics B will publish research on quantum physics and superconductivity, whereas *PMC Physics C* will cover biological and interdisciplinary physics. PhysMath Central plans to launch another four journals this year.

Venter reorganization closes genomics institute

The institution that produced the first complete genome sequence of a free-living organism has officially ceased to exist. Last week, the J. Craig Venter Institute (JCVI) in Rockville, Maryland, announced that 15 years after it was founded The Institute for Genomic Research — popularly known as TIGR — would be dissolved into the larger Venter Institute.

"We have long been leaders in genomics, and with our newly organized institute, I am certain we are poised to continue to blaze new trails in this field," said Venter, who is president and chairman of the JCVI, in a press release on 11 April. The newly reorganized, 500-staff institute will include research groups on plant genomics and microbial genomics, on which much of TIGR's work focused.

Claire Fraser-Liggett, Venter's ex-wife and the president and director of TIGR, will leave to found the new Institute of Genome Sciences at the University of Maryland School of Medicine in Baltimore.

P. G. TUTHILL

Red square at night

No Kremlin, no St Basil's, no Lenin's tomb: the centre of attention in the spectacular nebula that has been dubbed 'Red Square' is just a star called MWC 922. But it is possible that the square will, eventually, boast a mausoleum of its own, as that star may be headed towards death by supernova.

The diagonal lines caught in this image, which uses infrared as well as visible wavelengths, mark the edges of two cones of hot gas expanding from the star in a peculiarly even way. They have been captured through the use of adaptive optics at the Mount Palomar Observatory in California.

BUSINESS

Moment of reckoning

Fresh laws on the regulation of medicines are working their way through the US Congress — but will they strike the right balance between public safety and innovation? **Meredith Wadman** investigates.

Teddy Kennedy has a well-earned reputation as the last great American liberal. His very name should be enough to strike fear into the heart of any sharp-suited industry lobbyist. And as chairman of the Senate's health committee, which oversees the Food and Drug Administration (FDA), he is very much in the driving seat as Congress moves forward with significant drug-safety legislation.

So is the industry expecting a draconian clampdown on what it can or cannot do? Not quite, most industry-watchers say. As their ideas for legislation take shape, Kennedy and his colleagues in the new Democratic leadership in Congress face conflicting pressures.

On the one hand, the 2004 withdrawal of the painkiller Vioxx, after it caused heart problems in thousands of patients, has increased public pressure to tighten drug regulation. But on the other hand, the US drug and biotechnology industries — not least in Kennedy's home state of Massachusetts — are now mainstays of the US economy. They see a nimble and reasonably flexible drug-regulation system as crucial to their continued international success — and will bring all of their considerable lobbying influence to bear in pressing home the point.

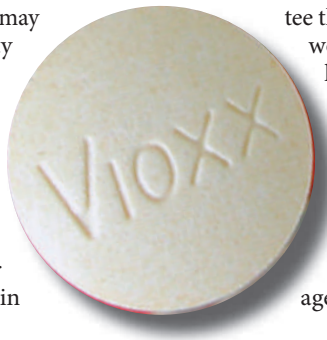
"We are hopeful that we will be able to work something through that the different stakeholders can support," Kennedy said last week as he dashed from the Capitol to his office nearby. "That is clearly our objective. But it is a complex issue. It is a continuing process."

Earlier, Kennedy had argued that his planned legislation will carefully balance different interests. "Our proposal includes a structure to oversee safety that is flexible enough to be tailored to the unique characteristics of each new drug, and strong enough to protect patients from unacceptable risk," he said.

The Vioxx affair and the Democrat takeover of Congress in January have pushed drug regulation up the legislative agenda. But the real reason that Congress may end up passing drug-safety laws this year, rather than just talking about them, is a technical one. The Prescription Drug User Fee Act, which directs more than US$300 million in fees from drug companies every year to pay for faster FDA drug reviews, expires in September. If that money is to keep flowing — as it must — the legislation needs to pass. And that provides lawmakers with a handy vehicle for changing drug-safety rules.

Ted Kennedy claims that his proposed legislation will work for regulators, patients and industry alike.

C. KRUPA/AP



M. DAY MASSEY/ZUMA/CORBIS

Kennedy and his Republican counterpart, Mike Enzi of Wyoming, are capitalizing on the momentum to renew the user-fee law. They want to fold into it a drug-safety bill that, drawing on a report by the Institute of Medicine (see *Nature* **443,** 372; 2006), boosts FDA enforcement muscle while requiring companies to submit strategies that assess the serious risks of every new drug. Companies will also need to set out and justify their plans to identify unexpected risks once a drug is on the market.

In a draft bill introduced in February and due to be refined by Kennedy's committee this week, the Kennedy–Enzi bill would allow the FDA to demand label changes on drugs when new safety concerns arise, rather than negotiating with companies as it does now. It would also give the agency power to forbid direct-to-consumer advertising of selected new drugs for up to two years.

And the law would allow the agency, at its discretion, to restrict distribution of high-risk drugs to specific locations, such as hospitals, or to particular providers, such as specially certified doctors.

Drugmakers could be held responsible for ensuring that the restricted distribution is working and, if not, for making sure that doctors and hospitals that do not comply no longer receive the restricted drug. They could also be required to conduct and publish fully fledged, post-approval clinical trials. A similar bill was introduced in the House last month, authored by Henry Waxman (Democrat, California) and Ed Markey (Democrat, Massachusetts).

Drug companies say that the Kennedy–Enzi bill goes too far. "This bill would tighten up the drug-safety requirements beyond a reasonable level," says Scott Lassman, a senior lawyer at the Pharmaceutical Research and Manufacturers of America (PhRMA), the drug industry's lobbying group in Washington DC.

The Biotechnology Industry Organization hasn't commented on the particulars of the draft, but has objected to the idea of using the user-fee act as a means of passing new drug-safety rules. The user-fee act "should not be encumbered or slowed down by any unrelated or scientifically contentious issues," says the organization's president Jim Greenwood.

Bruce Burlington, a vice-president of New Jersey drugmaker Wyeth who spent 17 years in senior positions at the FDA, told Kennedy at a health committee hearing on 14 March that his bill was a "thoughtful effort" to balance the twin goals of safety and providing patients with access to new drugs. Still, he said, the law should be less prescriptive, laying out general principles instead of requiring one-size-fits-all plans for risk management.

After the hearing, Burlington said that it is hard to gauge whether the bill would stifle innovation because it is written so broadly that much will depend on how it is administered. Nonetheless, he asserted, it "may end up freezing drugs out of development".

Another concern about the bill was voiced at the same hearing by Senator Tom Coburn (Republican, Oklahoma), a member of the health committee who is also a practicing obstetrician and gynaecologist. Coburn charged that as written, the bill would allow the FDA to forbid doctors from prescribing high-risk medicines for conditions other than those for which they are approved — an interpretation that Kennedy disputes. "The problem I see is that the FDA is going to be interfering with the practice of medicine," Coburn said.

Corrupting influence

At the other end of the political spectrum, a group of 21 academics and clinicians led by former FDA official Susan Wood of George Washington University in Washington DC is calling on Congress to throw out the entire user-fee law, claiming that it corrupts the agency. In a letter to Congress, the group wrote that the existing law "has helped to foster the public's perception that industry has become the primary client of the FDA, rather than the American people." They are calling on Congress to fund the whole of the FDA directly. That is something few expect to happen, for cost reasons.

For its part, the drug industry argues that the FDA just needs the resources to use the muscle it already has. Industry lobbyists note that the FDA's own recommendations for the new user-fee bill — submitted last month — would triple the user fees now directed to post-market safety, to $150 million over five years. That would allow the agency to hire 82 new employees. Another, entirely new user fee would allow it to hire 27 workers to vet direct-to-consumer drug advertisements before they run.

The PhRMA agreed to fund this expansion in closed-door negotiations with the regulator last year. "We think it provides the FDA with pretty much everything they need to address drug safety," says Lassman. The legislation that must pass by September will reveal whether Congress agrees. ■

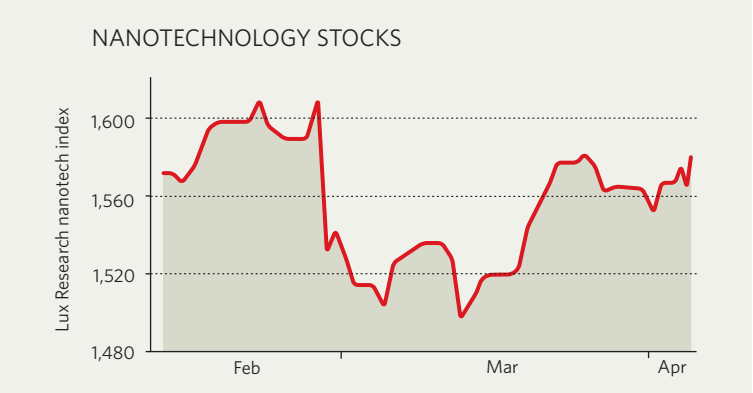
IN BRIEF

MULTIPLE WAFER **IBM says that it has developed a three-dimensional silicon chip. The computer company said that its technology, which uses tiny metal wires to connect silicon chips tightly stacked on top of each other, will enable it to pack more functionality into its products. A device based on the concept will debut in power chips in some of IBM's wireless devices later this year, and then spread to more general application.**

STOCK-MARKET GOAL **Russian drug company Pharmstandard is planning to float on the London and Moscow stock exchanges in June. The company sells generic drugs and is one of Russia's leading pharmaceutical companies. It has been valued by some analysts at around US$1.5 billion, and it will be the largest Russian drug company so far to float on international markets. Pharmstandard, which is part-owned by Roman Abramovich, the billionaire owner of Chelsea Football Club, has annual sales of about $320 million in the fast-growing Russian pharmaceutical market.**

REAGENT MERGER **Agilent Technologies, a manufacturer of scientific instruments based in Santa Clara, California, said that it would buy the biotechnology company Stratagene for around $246 million. Stratagene, which is based in La Jolla, California, develops clinical diagnostics and reagents and will be incorporated into Agilent's life-sciences unit if, as expected, shareholders approve the deal. The acquisition is set to boost Agilent's standing in the growing molecular-diagnostics market. Stratagene's stock rose $2.20 to $10.71 after the deal was announced on 6 April, and Agilent's stock went up 35 cents to $35.06.**

MARKET WATCH



Major stock indices got the jitters in late February over the short-term outlook for the US economy. But nanotechnology stocks, as measured by the Lux Research nanotech Index, held up quite well — and some companies had substantial gains.

Much of the general market concern is over a possible weakness in consumer demand if house prices in the United States continue to stagnate. But revenue for nanotechnology companies remains dominated by demand from businesses (in equipment and materials, for example), rather than from consumers, explains Peter Hébert, chief executive of New York-based Lux Research, which compiles the index. So the index has fared reasonably well, outperforming the high-tech Nasdaq index so far this year.

There has been no clear, general trend in the index over the past two months, but a few star stocks have performed well.

Arrowhead of Pasadena, California, for example, pleased its investors with several acquisitions, including that of Carbon Nanotechnologies, the carbon nanotubes company founded by late Nobel laureate Richard Smalley. Arrowhead's shares soared from US$3.75 in mid-March to $4.95 last week.

Los Angeles drug-delivery specialist Abraxis Bioscience, whose main product is a nanoparticle-borne breast-cancer drug Abraxane, also rose in value after announcing 2006 sales of $765 million, up from $520 million the previous year.

Hébert says that overall, confidence in the nanotechnology sector remains strong, and has been driven by growing global investment during 2006 — $6.4 billion in public research, $5.3 billion in corporate research and development and $650 million in venture capital. ■

Colin Macilwain

Bruce Burlington, a vice-president of New Jersey drugmaker Wyeth who spent 17 years in senior positions at the FDA, told Kennedy at a health committee hearing on 14 March that his bill was a "thoughtful effort" to balance the twin goals of safety and providing patients with access to new drugs. Still, he said, the law should be less prescriptive, laying out general principles instead of requiring one-size-fits-all plans for risk management.

After the hearing, Burlington said that it is hard to gauge whether the bill would stifle innovation because it is written so broadly that much will depend on how it is administered. Nonetheless, he asserted, it "may end up freezing drugs out of development".

Another concern about the bill was voiced at the same hearing by Senator Tom Coburn (Republican, Oklahoma), a member of the health committee who is also a practicing obstetrician and gynaecologist. Coburn charged that as written, the bill would allow the FDA to forbid doctors from prescribing high-risk medicines for conditions other than those for which they are approved — an interpretation that Kennedy disputes. "The problem I see is that the FDA is going to be interfering with the practice of medicine," Coburn said.

Corrupting influence

At the other end of the political spectrum, a group of 21 academics and clinicians led by former FDA official Susan Wood of George Washington University in Washington DC is calling on Congress to throw out the entire user-fee law, claiming that it corrupts the agency. In a letter to Congress, the group wrote that the existing law "has helped to foster the public's perception that industry has become the primary client of the FDA, rather than the American people." They are calling on Congress to fund the whole of the FDA directly. That is something few expect to happen, for cost reasons.

For its part, the drug industry argues that the FDA just needs the resources to use the muscle it already has. Industry lobbyists note that the FDA's own recommendations for the new user-fee bill — submitted last month — would triple the user fees now directed to post-market safety, to $150 million over five years. That would allow the agency to hire 82 new employees. Another, entirely new user fee would allow it to hire 27 workers to vet direct-to-consumer drug advertisements before they run.

The PhRMA agreed to fund this expansion in closed-door negotiations with the regulator last year. "We think it provides the FDA with pretty much everything they need to address drug safety," says Lassman. The legislation that must pass by September will reveal whether Congress agrees. ■

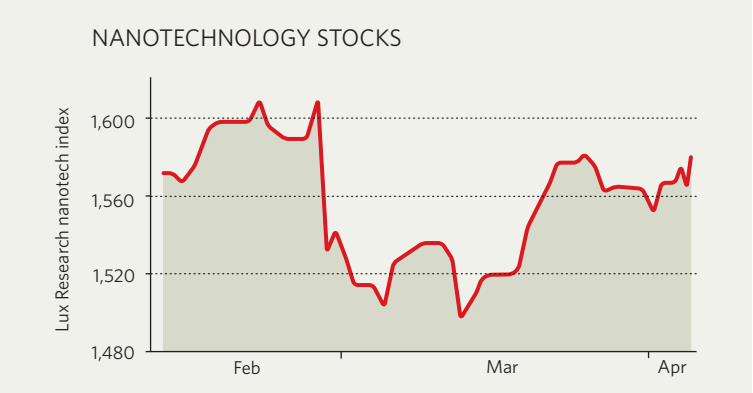
IN BRIEF

MULTIPLE WAFER **IBM says that it has developed a three-dimensional silicon chip. The computer company said that its technology, which uses tiny metal wires to connect silicon chips tightly stacked on top of each other, will enable it to pack more functionality into its products. A device based on the concept will debut in power chips in some of IBM's wireless devices later this year, and then spread to more general application.**

STOCK-MARKET GOAL **Russian drug company Pharmstandard is planning to float on the London and Moscow stock exchanges in June. The company sells generic drugs and is one of Russia's leading pharmaceutical companies. It has been valued by some analysts at around US$1.5 billion, and it will be the largest Russian drug company so far to float on international markets. Pharmstandard, which is part-owned by Roman Abramovich, the billionaire owner of Chelsea Football Club, has annual sales of about $320 million in the fast-growing Russian pharmaceutical market.**

REAGENT MERGER **Agilent Technologies, a manufacturer of scientific instruments based in Santa Clara, California, said that it would buy the biotechnology company Stratagene for around $246 million. Stratagene, which is based in La Jolla, California, develops clinical diagnostics and reagents and will be incorporated into Agilent's life-sciences unit if, as expected, shareholders approve the deal. The acquisition is set to boost Agilent's standing in the growing molecular-diagnostics market. Stratagene's stock rose $2.20 to $10.71 after the deal was announced on 6 April, and Agilent's stock went up 35 cents to $35.06.**

MARKET WATCH



Major stock indices got the jitters in late February over the short-term outlook for the US economy. But nanotechnology stocks, as measured by the Lux Research nanotech Index, held up quite well — and some companies had substantial gains.

Much of the general market concern is over a possible weakness in consumer demand if house prices in the United States continue to stagnate. But revenue for nanotechnology companies remains dominated by demand from businesses (in equipment and materials, for example), rather than from consumers, explains Peter Hébert, chief executive of New York-based Lux Research, which compiles the index. So the index has fared reasonably well, outperforming the high-tech Nasdaq index so far this year.

There has been no clear, general trend in the index over the past two months, but a few star stocks have performed well.

Arrowhead of Pasadena, California, for example, pleased its investors with several acquisitions, including that of Carbon Nanotechnologies, the carbon nanotubes company founded by late Nobel laureate Richard Smalley. Arrowhead's shares soared from US$3.75 in mid-March to $4.95 last week.

Los Angeles drug-delivery specialist Abraxis Bioscience, whose main product is a nanoparticle-borne breast-cancer drug Abraxane, also rose in value after announcing 2006 sales of $765 million, up from $520 million the previous year.

Hébert says that overall, confidence in the nanotechnology sector remains strong, and has been driven by growing global investment during 2006 — $6.4 billion in public research, $5.3 billion in corporate research and development and $650 million in venture capital. ■

Colin Macilwain

Bruce Burlington, a vice-president of New Jersey drugmaker Wyeth who spent 17 years in senior positions at the FDA, told Kennedy at a health committee hearing on 14 March that his bill was a "thoughtful effort" to balance the twin goals of safety and providing patients with access to new drugs. Still, he said, the law should be less prescriptive, laying out general principles instead of requiring one-size-fits-all plans for risk management.

After the hearing, Burlington said that it is hard to gauge whether the bill would stifle innovation because it is written so broadly that much will depend on how it is administered. Nonetheless, he asserted, it "may end up freezing drugs out of development".

Another concern about the bill was voiced at the same hearing by Senator Tom Coburn (Republican, Oklahoma), a member of the health committee who is also a practicing obstetrician and gynaecologist. Coburn charged that as written, the bill would allow the FDA to forbid doctors from prescribing high-risk medicines for conditions other than those for which they are approved — an interpretation that Kennedy disputes. "The problem I see is that the FDA is going to be interfering with the practice of medicine," Coburn said.

Corrupting influence

At the other end of the political spectrum, a group of 21 academics and clinicians led by former FDA official Susan Wood of George Washington University in Washington DC is calling on Congress to throw out the entire user-fee law, claiming that it corrupts the agency. In a letter to Congress, the group wrote that the existing law "has helped to foster the public's perception that industry has become the primary client of the FDA, rather than the American people." They are calling on Congress to fund the whole of the FDA directly. That is something few expect to happen, for cost reasons.

For its part, the drug industry argues that the FDA just needs the resources to use the muscle it already has. Industry lobbyists note that the FDA's own recommendations for the new user-fee bill — submitted last month — would triple the user fees now directed to post-market safety, to $150 million over five years. That would allow the agency to hire 82 new employees. Another, entirely new user fee would allow it to hire 27 workers to vet direct-to-consumer drug advertisements before they run.

The PhRMA agreed to fund this expansion in closed-door negotiations with the regulator last year. "We think it provides the FDA with pretty much everything they need to address drug safety," says Lassman. The legislation that must pass by September will reveal whether Congress agrees. ■

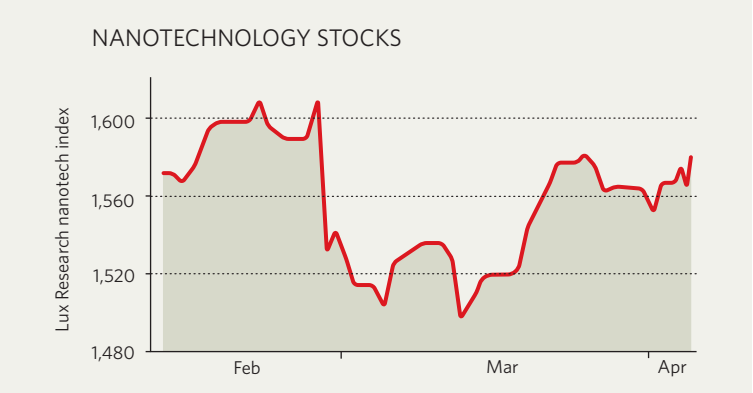
IN BRIEF

MULTIPLE WAFER **IBM says that it has developed a three-dimensional silicon chip. The computer company said that its technology, which uses tiny metal wires to connect silicon chips tightly stacked on top of each other, will enable it to pack more functionality into its products. A device based on the concept will debut in power chips in some of IBM's wireless devices later this year, and then spread to more general application.**

STOCK-MARKET GOAL **Russian drug company Pharmstandard is planning to float on the London and Moscow stock exchanges in June. The company sells generic drugs and is one of Russia's leading pharmaceutical companies. It has been valued by some analysts at around US$1.5 billion, and it will be the largest Russian drug company so far to float on international markets. Pharmstandard, which is part-owned by Roman Abramovich, the billionaire owner of Chelsea Football Club, has annual sales of about $320 million in the fast-growing Russian pharmaceutical market.**

REAGENT MERGER **Agilent Technologies, a manufacturer of scientific instruments based in Santa Clara, California, said that it would buy the biotechnology company Stratagene for around $246 million. Stratagene, which is based in La Jolla, California, develops clinical diagnostics and reagents and will be incorporated into Agilent's life-sciences unit if, as expected, shareholders approve the deal. The acquisition is set to boost Agilent's standing in the growing molecular-diagnostics market. Stratagene's stock rose $2.20 to $10.71 after the deal was announced on 6 April, and Agilent's stock went up 35 cents to $35.06.**

MARKET WATCH



Major stock indices got the jitters in late February over the short-term outlook for the US economy. But nanotechnology stocks, as measured by the Lux Research nanotech Index, held up quite well — and some companies had substantial gains.

Much of the general market concern is over a possible weakness in consumer demand if house prices in the United States continue to stagnate. But revenue for nanotechnology companies remains dominated by demand from businesses (in equipment and materials, for example), rather than from consumers, explains Peter Hébert, chief executive of New York-based Lux Research, which compiles the index. So the index has fared reasonably well, outperforming the high-tech Nasdaq index so far this year.

There has been no clear, general trend in the index over the past two months, but a few star stocks have performed well.

Arrowhead of Pasadena, California, for example, pleased its investors with several acquisitions, including that of Carbon Nanotechnologies, the carbon nanotubes company founded by late Nobel laureate Richard Smalley. Arrowhead's shares soared from US$3.75 in mid-March to $4.95 last week.

Los Angeles drug-delivery specialist Abraxis Bioscience, whose main product is a nanoparticle-borne breast-cancer drug Abraxane, also rose in value after announcing 2006 sales of $765 million, up from $520 million the previous year.

Hébert says that overall, confidence in the nanotechnology sector remains strong, and has been driven by growing global investment during 2006 — $6.4 billion in public research, $5.3 billion in corporate research and development and $650 million in venture capital. ■

Colin Macilwain

THE CANDIDATES RESPOND

ALAMY; E. FEFERBERG/AP; P. KOVARIK/AP; G. BASSIGNAC/GAMMA, CAMERA PRESS LONDON

France's presidential elections are taking place at a time of deep debate over the French research community's standing and prospects. To further the debate, *Nature*'s **Declan Butler** submitted a list of questions on research issues to the three leading candidates. Their full responses, in both French and English, are on our website. Here we present extracts:

Is French science in decline and, if so, why? How much will you invest in science, when and on what priorities?

Nicolas Sarkozy: It's true that we've had various warning signs over the past few years that the relative position of French science in the world is being eroded. France nonetheless maintains expertise of the highest international level in many disciplines, in particular in mathematics, physics and engineering. I note too that France exports its scientific expertise abroad, even if I regret the fact that many of our young scientists increasingly choose to leave the country because they no longer feel they can succeed at home.

Research and higher education will be at the forefront of my priorities. Although this will be realized in the shape of more resources — €4 billion [US$5.4 billion] extra for research, and €5 billion for higher education — it will also involve deep reforms in the way the system works. Reform without resources would be as fruitless as resources without reform.

I want to favour powerful and autonomous universities, which will be reinstated at the core of our research effort, and to reinforce a culture of scientific evaluation by promoting competitive grants. I would also note that the reshaping and restructuring of the Saclay plateau will be a major presidential initiative. This site is unique in Europe in its enormous concentration of centres of education and research, but there is scope for it to be exploited yet further.

Ségolène Royal: I don't think it's fair to speak of a decline, because France still has great assets and renowned university researchers, for example in mathematics. But for years, France has not really made research or the universities a top priority. The recruitment drive launched by Lionel Jospin was stopped when the conservatives returned to power; budgets have stagnated, PhD students and young scientists have been neglected, and the research system has been made more complex. France has fallen from 5th to 11th place among OECD countries in terms of science spending. I want to turn that situation around by making higher education, research and innovation top spending priorities, with a 10% budget increase each year for five years.

François Bayrou: It's not so much that French science is in decline, but rather that other countries, including the United States, Japan, the nations of northern Europe and, more recently, China, have constantly increased investment in research over the years. For too long, research budgets in France have stagnated. Despite this situation, French scientists still rank among

Who's who

Nicolas Sarkozy (pictured centre) is head of the right-of-centre UMP party, which was founded by outgoing president Jacques Chirac. He has served as both finance minister and minister of the interior.

Ségolène Royal (pictured left) is the Socialist Party candidate. A member of the National Assembly, she is president of the Poitou-Charentes region in the west of France.

François Bayrou (pictured right) is the head of a small centrist party, the UDF. He was minister for education in two right-wing governments during the 1990s, and has been a member of the European Parliament.

There are also nine other candidates, including **Jean-Marie Le Pen** of the National Front, **Dominique Voynet** of the Greens, **Olivier Besancenot** of the Revolutionary Communists, **Philippe de Villiers** of the traditionalist and eurosceptic Movement for France, and anti-globalization campaigner **José Bové**.

The two candidates who get the most votes in the first round of the elections, which in metropolitan France will be held on 22 April, will proceed to a run-off election on the first weekend in May.

Issues and initials

A guide to what the candidates and scientists are talking about.

ANR A national research-funding agency founded in 2005. Previously, most research funding in France had been via block grants to specific research agencies.

Ariane Europe's family of space launch vehicles, built by a consortium led by French industry and a source of great French pride.

'Competitiveness clusters' A 2005 interministerial initiative to link companies, research and education centres.

CNRS France, and Europe's, largest basic-research organization

EPR European pressurized reactor, a new type of nuclear plant designed by Areva — a French engineering company in which Germany's Siemens has a large stake — and the French electricity utility Electricité de France.

Galileo An EU/European Space Agency programme of positioning satellites that would rival or complement (take your pick) the American Global Positioning System.

GMES Global monitoring for environment and security, a joint EU/European Space Agency programme with some military ramifications.

GMOs Genetically modified organisms. A source of considerable unease in parts of rural France, and a sore point with minor presidential candidate José Bové.

Grandes Ecoles The actually quite small schools that form the élite branch of French higher education, producing just a few thousand graduates every year; almost every school in the system is focused on a single speciality.

INRA The national agricultural research organization.

INSERM The national medical research organization.

PhD Doctorates are a hot-button issue in France, where a PhD was until recently not recognized as a professional qualification (it still isn't under French law), and a poor cousin of a diploma from a Grande Ecole.

PRES Research and higher-education hubs. New entities intended to bring together nearby universities, hospitals and public and private research organizations, created by the 2006 law on research.

Saclay A site outside Paris that is home to the Soleil synchrotron, ten research organizations and ten higher-education centres including three Grandes Ecoles.

the best in many disciplines, in particular in mathematics, and in certain fields in physics, chemistry, biology, social sciences and the environment. It is now urgent to invest in emerging disciplines. I propose that France reinvest in research, with a 5% budget increase in real terms every year for ten years.

Is French science beyond reform?

Mme Royal: The real question is: which reforms does French science need? It's our misfortune that successive right-wing governments have failed to make research a priority. It was opposition to the impoverishment of the university and its decline that brought students out on to the streets in 1986 and 1995, and scientists in 2004.

M. Bayrou: That's not fair to French research. I am very conscious of its willingness to get beyond the deadlock in which it finds itself. Research must first be reinstated as a priority on the political agenda; leading French politicians must make that their personal duty.

M. Sarkozy: No, French science is not unreformable! I think it has never been more ripe for reform. Growing awareness of the state of research is spurring demand for change. Lecturers and researchers are faced with globalization and international competition, just as everyone else is. Travelling a lot in their jobs, they are able to make comparisons.

The leadership of France's industries and government is dominated by graduates not from universities, but from the Grandes Ecoles. How would you reconcile the parts of this dual system so as to recognize the value of PhD scientists?

M. Bayrou: Rather than pitting the different educational routes against each other, I prefer policies that diminish the differences between them. For example, we might extend the existing bridges between universities and the Grandes Ecoles, such as reorientation classes, to deserving students in all courses at the master's or PhD level; or it might be possible for the Grandes Ecoles and universities to set up common curricula.

M. Sarkozy: Research students will come into their own when French universities finally have available the finances and the autonomy they need to be centres of excellence — something that is already the case for courses in law, medicine and economics.

Research needs to play a bigger role at the Grandes Ecoles, and the best university students should be able to switch to those courses. Universities and Grandes Ecoles that are close to each other should have joint campuses with shared services. Universities could benefit from the Grandes Ecoles' business know-how, and the access their students enjoy to highly responsible jobs. I want PhD students to be able to access opportunities beyond the areas of research and education.

Mme Royal: I have pledged to have the PhD recognized both in the civil service and in the private sector. It is also necessary to bring the various parts of the higher-education system closer together, and to establish routes for moving between the Grandes Ecoles and the universities. That will be a role for the PRES, which will stimulate interaction and synergy between universities, Grandes Ecoles and research organizations.

How would you modernize France's universities?

M. Sarkozy: As of the day after the elections, I will be ready to launch a major reform of French universities designed to give them much more autonomy. This will include powers to recruit, to fix salaries, to decide how they organize themselves, to build endowments and to diversify their funding sources. I will also rebuild the way that they are governed, restructuring their executive boards and the ways they choose their presidents.

M. Bayrou: After a massive increase in student numbers over the past two decades, enrolment has now stabilized, and this makes it possible to envisage a new phase of long-term development. The universities suffer three ills: the absence of recognition of the PhD, lack of funding and a poorly adapted governance structure. We need to reach spending-per-student levels equal to or more than the average of OECD countries, continue the rapprochement with the Grandes Ecoles that has now begun, and make changes in the ways the universities are run.

Mme Royal: I favour a rational, optimal use of resources based on evaluation; this means we must provide favourable working conditions for all researchers. If we supported only a small proportion of researchers it would mean that we would be paying the others without benefiting from their potential. That would be absurd.

State planning created France's powerful aerospace, nuclear and transport industries, but is 'innovation by decree' possible for fast-moving sectors such as biotechnology and information technology?

M. Bayrou: Research is not in itself innovation, but it can and should contribute to it. There are several ways of promoting this. Researchers

The incoming French president will need to clarify the country's role in space exploration (right) as well as wrestle with emotive issues such as stem-cell therapies.

A. MARCHI/GAMMA, CAMERA PRESS LONDON

ESA/CNES/ARIANESPACE-SERVICE OPTIQUE CSG

contribute to our growth and our competitivity by taking interest in their work's potential for wealth creation. To attract talent, I propose that for researchers living in France, royalties from patents should be exempt from income tax. I want to reinforce technology-transfer departments in our research institutes — a few hundred million euros could have a substantial impact — to help the birth and development of innovative enterprises.

Mme Royal: Innovation has stagnated because fundamental research has not been supported. The proactive policies of the 1960s led to the development of the space, aeronautics and nuclear-power industries. The 'breakdown' in biotechnology — despite the fact that French labs were at the forefront of sequencing the human genome in the early 1990s — stems from a glaring lack of support for the life sciences a decade ago.

M. Sarkozy: The competitiveness clusters that I established in my different positions in government are still very young. But an initial appraisal is rather encouraging. I would emphasize that they represent a relatively innovative approach, at least by French standards. As you say in Great Britain it is a bottom-up, and not top-down, approach.

It is, above all, the quality of research and the dynamism of the ways in which it is disseminated that determine our potential for innovation — not just financial incentives. But let's be pragmatic; everywhere in the world, from emerging economies to developed and ostensibly free-market countries, the state intervenes to encourage innovation, and to build and reinforce the industrial and technological sectors of the future.

What cuts in greenhouse-gas emissions would you commit France to and how would you attain them? What should be agreed post-Kyoto?

Mme Royal: I am committed to a 75% reduction in our greenhouse-gas emissions by 2050. I'm setting out an ambitious policy because I'm convinced that climate change is the major challenge of the twenty-first century. So I advocate saving energy in the building and transport sectors, developing renewable energy (solar, wind, biomass and geothermal) and promoting research into carbon capture and storage, electricity storage, hydrogen, intelligent electricity grids and new fuels created through hydrolysis of biomass.

M. Sarkozy: I am proud to be able to say that France made a visionary choice in committing itself several decades ago to developing its nuclear-power programme. Just think, the carbon emissions saved by France using nuclear-power stations rather than fossil fuels are equivalent to those of all Europe's cars. Of course, we have also made scientific and technological priorities of research on renewable energy and more energy-efficient means of production and transport.

As a market of 500 million people, Europe should make greater efforts to encourage its large commercial partners — in particular the United States, China and Canada — to play according to the planet's own rules. Countries that behave like stowaways hitching a free ride, making no effort to reduce their emissions, should not continue to benefit from the competitive industrial advantage this gives them. To compensate for this we must tax products from countries that make no effort to reduce emissions after 2012, even if this means modifying World Trade Organization rules.

M. Bayrou: I will fulfil the European commitment to a target of generating 20% of our energy from renewable sources. But we need to do even more, through saving energy and new technologies — for example, the construction sector already has the know-how to erect zero-emission buildings. That will create local jobs.

Would you maintain nuclear power's current 75% share of French electricity generation? And how will you tackle France's accumulating nuclear waste?

M. Sarkozy: The nuclear sector is of absolute strategic importance, as well as of industrial and technological excellence. France is one of the rare countries to have mastered the nuclear-fuel cycle in its entirety. France will continue to nurture its comparative advantage here by modernizing its nuclear fleet and know-how. That's why we have committed to a series of third-generation reactors, the EPR, and a research programme into fourth-generation reactors.

Mme Royal: I will pay particular attention to guaranteeing that the storage of nuclear waste is reversible. Parliament will decide in ten years time which options should be retained for long-term management of nuclear waste.

I think that the current government took the decision to go ahead with the EPR without adequate analysis or debate. No impact assessment was presented to parliament, and no effort was made to create real diversification in our energy mix. We cannot set our country's energy future in stone without an in-depth debate, not just on the EPR but on the entire issue.

M. Bayrou: The EPR project will be maintained. But we need a scientific assessment of it. The renewal of our existing fleet of reactors hinges on this project, and a decision this important for our energy policy cannot be taken on the sly. There must be as wide and democratic a debate as possible.

A demonstration project should be launched rapidly to prove that after temporary storage, the volume and radioactivity of waste can be reduced to low levels. This is necessary to reassure the many men and women in France, and worldwide, who have doubts about this form of energy because of the risks still associated with it.

What are your priorities for space?

Mme Royal: France is a major space power, having developed, with its European partners, the high-quality Ariane launchers. I think that the high costs of manned space flight mean that such ventures should be carried out in international programmes. The launch of Earth-observation satellites is, of course, a priority.

M. Bayrou: The emphasis should also be on fundamental research and exploration of the Universe, a field in which France and Europe are proficient, and then on GMES and on the Galileo navigation system. For Europe to carry sufficient weight, should it aim for the Moon or even farther, perhaps going it alone? Should it support the International Space Station? This all demands reflection.

M. Sarkozy: The Galileo project has become bogged down in national quarrels that are petty compared with the stakes on the table, and breaking the current deadlock is a matter of urgency.

I'm keen on greater European cooperation in space. France should be ready to make the extra effort to lead the way, if needs be, as it has done in the past. The main goal must be to maintain and reinforce our basic civil, military and scientific skills. If, after that, we can together develop more ambitious manned flight and planetary-exploration missions, then why not?

What will be your policy on genetically modified crops? Would you change existing laws on embryonic stem-cell research?

M. Bayrou: What has happened with GMOs is symptomatic of the lack of democratic consultation on major topics in France. Such a debate would have helped to bring out the citizens' expectations, and to provide directions for research, so as to allow the downstream use of the technology in a way that was regulated and acceptable. We must therefore organize this now, with input from independent scientists. In the meantime, I am in favour of an immediate moratorium on GMOs.

M. Sarkozy: I want research to continue on GMOs as, among other things, that is the only way to improve our knowledge of the potential risks to human health and biodiversity. Unless we can be highly certain that they are harmless, I am less enthusiastic about their industrialization and marketing.

As for research on embryonic stem cells, with the creation of the Agency of Biomedicine in 2005, researchers can now submit research projects for review. I'm delighted that projects that have been favourably reviewed can thus develop in our country within a clear and evolving framework. I think it's indispensable that research also continues to develop in this area.

Mme Royal: I am in favour of a moratorium on open-field cultivation of GMOs and of having a public debate on this question, which is of interest to all citizens.

Research on human stem cells should be permitted provided that they are obtained after informed consent, they come from embryos that are no longer part of any fertility treatments and the proposed protocol has been rigorously examined. We must revise the legal framework to reconcile ethical principles and scientific progress. ■

For fuller answers, and for questions on the street protests of 2004, nuclear deterrence, the CNRS, EU science and the common agricultural policy, visit our website at http://tinyurl.com/23fwj5.
See also Editorial, page 831.

Let science speak for itself

You've heard what the presidential candidates think the challenges facing science in France are. *Nature* also canvassed opinion across the French research spectrum: from young researchers to reformers and industrialists. **Declan Butler** reports.

Pierre Chambon

Biologist at the Institute of Genetics and Molecular and Cellular Biology in Illkirch near Strasbourg, of which he was formerly director.

The common assertion that French research is doing badly is untrue; it's doing badly in some sectors, with life sciences the main concern. The French research system hasn't been organized in 40 years. There's duplication in biology across the CNRS, INSERM, INRA and the universities that's damaging competitiveness. We need a single life-sciences agency, plus perhaps a second for more applied work, to bring together the best life scientists and give them the means to do top-quality research.

Biotech companies and universities compete for the best scientists worldwide. But France's civil-service pay scales mean that a biologist earns the same as a sociologist or anthropologist. This makes it impossible to attract the world's best biologists, but life sciences in France cannot survive on French researchers alone.

In France, most scientists start and finish their careers in the same organization, whether this is a research agency or a university. There are two separate corps of scientists, some teaching and researching, some doing only research. This is inefficient, because you want less effective researchers doing more teaching, and the best researchers doing less, according to their success in winning grants. For this to work, we need to have just a single corps of researchers, all attached to universities, with the research agencies transformed into research councils.

Mme Royal: I will pay particular attention to guaranteeing that the storage of nuclear waste is reversible. Parliament will decide in ten years time which options should be retained for long-term management of nuclear waste.

I think that the current government took the decision to go ahead with the EPR without adequate analysis or debate. No impact assessment was presented to parliament, and no effort was made to create real diversification in our energy mix. We cannot set our country's energy future in stone without an in-depth debate, not just on the EPR but on the entire issue.

M. Bayrou: The EPR project will be maintained. But we need a scientific assessment of it. The renewal of our existing fleet of reactors hinges on this project, and a decision this important for our energy policy cannot be taken on the sly. There must be as wide and democratic a debate as possible.

A demonstration project should be launched rapidly to prove that after temporary storage, the volume and radioactivity of waste can be reduced to low levels. This is necessary to reassure the many men and women in France, and worldwide, who have doubts about this form of energy because of the risks still associated with it.

What are your priorities for space?

Mme Royal: France is a major space power, having developed, with its European partners, the high-quality Ariane launchers. I think that the high costs of manned space flight mean that such ventures should be carried out in international programmes. The launch of Earth-observation satellites is, of course, a priority.

M. Bayrou: The emphasis should also be on fundamental research and exploration of the Universe, a field in which France and Europe are proficient, and then on GMES and on the Galileo navigation system. For Europe to carry sufficient weight, should it aim for the Moon or even farther, perhaps going it alone? Should it support the International Space Station? This all demands reflection.

M. Sarkozy: The Galileo project has become bogged down in national quarrels that are petty compared with the stakes on the table, and breaking the current deadlock is a matter of urgency.

I'm keen on greater European cooperation in space. France should be ready to make the extra effort to lead the way, if needs be, as it has done in the past. The main goal must be to maintain and reinforce our basic civil, military and scientific skills. If, after that, we can together develop more ambitious manned flight and planetary-exploration missions, then why not?

What will be your policy on genetically modified crops? Would you change existing laws on embryonic stem-cell research?

M. Bayrou: What has happened with GMOs is symptomatic of the lack of democratic consultation on major topics in France. Such a debate would have helped to bring out the citizens' expectations, and to provide directions for research, so as to allow the downstream use of the technology in a way that was regulated and acceptable. We must therefore organize this now, with input from independent scientists. In the meantime, I am in favour of an immediate moratorium on GMOs.

M. Sarkozy: I want research to continue on GMOs as, among other things, that is the only way to improve our knowledge of the potential risks to human health and biodiversity. Unless we can be highly certain that they are harmless, I am less enthusiastic about their industrialization and marketing.

As for research on embryonic stem cells, with the creation of the Agency of Biomedicine in 2005, researchers can now submit research projects for review. I'm delighted that projects that have been favourably reviewed can thus develop in our country within a clear and evolving framework. I think it's indispensable that research also continues to develop in this area.

Mme Royal: I am in favour of a moratorium on open-field cultivation of GMOs and of having a public debate on this question, which is of interest to all citizens.

Research on human stem cells should be permitted provided that they are obtained after informed consent, they come from embryos that are no longer part of any fertility treatments and the proposed protocol has been rigorously examined. We must revise the legal framework to reconcile ethical principles and scientific progress. ■

For fuller answers, and for questions on the street protests of 2004, nuclear deterrence, the CNRS, EU science and the common agricultural policy, visit our website at http://tinyurl.com/23fwj5.
See also Editorial, page 831.

Let science speak for itself

You've heard what the presidential candidates think the challenges facing science in France are. *Nature* also canvassed opinion across the French research spectrum: from young researchers to reformers and industrialists. **Declan Butler** reports.

Pierre Chambon

Biologist at the Institute of Genetics and Molecular and Cellular Biology in Illkirch near Strasbourg, of which he was formerly director.

The common assertion that French research is doing badly is untrue; it's doing badly in some sectors, with life sciences the main concern. The French research system hasn't been organized in 40 years. There's duplication in biology across the CNRS, INSERM, INRA and the universities that's damaging competitiveness. We need a single life-sciences agency, plus perhaps a second for more applied work, to bring together the best life scientists and give them the means to do top-quality research.

Biotech companies and universities compete for the best scientists worldwide. But France's civil-service pay scales mean that a biologist earns the same as a sociologist or anthropologist. This makes it impossible to attract the world's best biologists, but life sciences in France cannot survive on French researchers alone.

In France, most scientists start and finish their careers in the same organization, whether this is a research agency or a university. There are two separate corps of scientists, some teaching and researching, some doing only research. This is inefficient, because you want less effective researchers doing more teaching, and the best researchers doing less, according to their success in winning grants. For this to work, we need to have just a single corps of researchers, all attached to universities, with the research agencies transformed into research councils.

French researchers took to the streets in 2004 in protest at what they saw as the government's attitude of neglect towards science.

E. DESSONS/JDD/GAMMA, CAMERA PRESS LONDON

The efficiency of French research won't improve without more flexibility about who gets what during their career. France wants top research but rejects élitism; that's not possible.

Biology also lacks a strong industrial lobby in France to push for greater public support. When the US National Institutes of Health's budget doubled in [the] five years [to 2003], France's was flat. The French drug industry is weak, comprising mainly Sanofi-Aventis [a multinational with much of its research in the United States] and a few other companies.

In Strasbourg, our institute was the first in Europe for molecular geneticists, and in the top 10 worldwide. But soon we will have lost all our top foreign scientists. This is not just because of the salaries; people would prefer to stay in France because of the quality of life. But they feel they are wasting their time having to look right and left for scraps. They can get better working conditions, and salaries, abroad.

There's no point in creating more posts unless people are given the means to do top-quality research, and there's no point in increasing funding if we don't reform our structures, otherwise it is like watering sand. Unless we attack the problems of the research agencies, modernizing French research will be mission impossible. We, the researchers, want reform.

Edouard Brézin
Physicist at the Ecole Normale Supérieure in Paris and ex-president of CNRS and the French Academy of Sciences.

The low attractiveness of science and engineering careers is a big problem in France. The young are increasingly turning to other careers. Moreover, the best students aim not to go to university or to get a PhD, but to graduate as an engineer from the élite Grandes Ecoles, as this offers better career prospects. So although at present French research is strong in many fields, I'm not at all sure that this will still be the case in 10–15 years time.

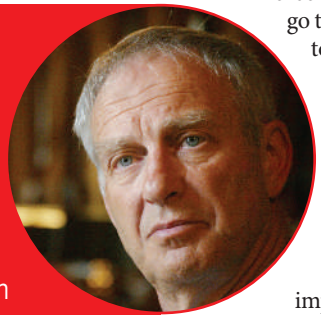
The high teaching loads imposed on young scientists starting work at a university prevent them from doing research during the very years when they should be cutting their research teeth. They are recruited on the basis of research talents, and we then annihilate their chances of establishing themselves. New recruits need to be freed up for research.

France has a chronic problem building links between public and private research. The concept of wealth creation in public research is perhaps less naturally engrained in our culture than in the United States and Britain. But I'm also struck by the fact that in French industry those in management rarely have any research training, as most are graduates of the Grandes Ecoles. Companies also tend to have a negative view of people with a research background — an engineer from one of the Grandes Ecoles who also has a PhD might be best advised to hide that fact in job interviews.

The relationship of universities to the Grandes Ecoles is one of many important questions facing the university system. There are important issues that need to be tackled to reform university governance, and to give them greater autonomy than they have in today's homogenized national system. But there hasn't been a major reform of French universities for decades, so all of these questions are off the radar.

The new ANR has its pros and cons. It's beneficial in that it makes it easier for young

"France has a chronic problem building links between public and private research. In French industry those in management rarely have any research training."
— Edouard Brézin



J. SAGET/AFP/GETTY

D. MONNIAUX

The Grande Ecoles, such as the ENS in Paris, have come to dominate French higher education.

research teams to attain funding, but the central government still decides research priorities and nominates management teams. Compare this with how its German counterpart, the DFG, is run; there the management is largely elected by the scientific community on the grounds of their scientific expertise. They are two different worlds.

That said, the French scientific community also needs to reform itself. The research agencies and universities elect various committees and evaluation panels, and I've long fought with the community for people to be elected on merit, and not on trade-union criteria. When I receive a voting bulletin, it doesn't give any indication of whether the candidate has even published in the past 5 years, just their trade-union profession of faith. It's incredible, ridiculous: a system from the Middle Ages.

Maïwenn Corrignan

President of the Young Researchers' Confederation and a sociologist at the University of Rennes 2 in Brittany.

"Many scientists would prefer not to stay at one research organization for their entire career, but there are too few bridges to allow greater mobility."
— Maïwenn Corrignan

Last year's decree that the PhD is a professional qualification was a real advance for young scientists here. It put an end to the widespread practice — particularly by research charities — of paying postgrads a minimal salary without benefits (or no salary at all). This change forces research agencies to provide contracts and pay social security and pension benefits, and also means postgrads on short-term contracts subsequently qualify for unemployment benefit.

The financial uncertainty of PhD students and the lack of recognition of the value of a PhD in recruitment was one of the main factors behind the street demonstrations of 2004. We don't want a postdoc system. It won't work in France, which has a culture of full-time employment; recruiters have a negative view of scientists with multiple short-term contracts as perpetual students.

We need a change in mentalities. Many scientists would prefer not to stay at one research organization for their entire career, but there are too few bridges to allow greater mobility between research agencies, universities and industry.

Better career guidance is also needed, so that the young are aware of opportunities beyond the research agencies and universities, both in industry and outside research. Things are improving; more than 20% of PhDs now go into industry, and that trend is increasing. Companies are also beginning to recruit, and value, those with PhDs. The young aren't disinterested in research; it's the poor working conditions that put them off.

Alain Trautmann

Cell biologist at the Cochin Institute in Paris and an instigator and former spokesman for the movement 'Save Research' that led protests in 2004.

In 2004, the scientific community in France organized an emergency nationwide science summit, and proposed a rash of reforms. But the government ignored them, and pushed ahead with its own. We are not against reform, but want changes to be decided in consultation with the science community. Successive governments have instead tried to impose unworkable reforms.

The current government is trying to impose an American system on France, without taking our history into account. They act as if the CNRS and the Grandes Ecoles, the main elements of the French system, don't exist. The fact that the Grandes Ecoles divert the brightest young people away from science is catastrophic. We need to evolve the French system, taking such realities into account.

In 2004 we proposed that, to make science more attractive to the young, the government tie its generous tax relief to companies not just on the basis of their degree of investment in research but also according to the number of trained scientists with PhDs they recruited. This could create change, and would not cost the state any extra, but the government rejected it. Ideologically, it was opposed by the lobby of the Grandes Ecoles, whose engineering graduates dominate the state apparatus.

The government also wants to make the universities the centrepiece of the research system. But France is currently one of the worst-placed countries in terms of investment in its universities. The universities that perform well are those that have joint laboratories with the CNRS. I'm not for maintaining the CNRS just because it already exists, but it's naive to believe, as some in government do, that one can simply abolish the CNRS overnight. Doing so without first reforming and reinvesting in the universities would just make the system worse.

André Choulika

Chief executive of Cellectis, a biotech spin-off from the Pasteur Institute in Paris.

France's competitiveness in research and innovation has emerged as an election issue in the 2007 presidential campaign. Significant investment in the life sciences in 2005 — €2.4 billion [US$3.2 billion], or 25.7% of the 2005 civil research budget — has so far failed to deliver the anticipated socioeconomic benefits.

The money has not been spent optimally. The multiplication of publicly supported research

organizations and structures has resulted in a failure to concentrate funds behind ambitious projects, and no clear policy for increasing competitiveness. The creativity and talent of French researchers is undisputed, but the quality of the country's research has deteriorated in recent years.

The ANR is a step in the right direction for improving research, provided it is given transparent governance and true independence. There is a need for greater translational research, and a culture that recognizes and rewards those whose research is outside the mainstream and ahead of the curve.

It seems timely to recall De Gaulle's famous quote: "We have plenty of researchers who seek, what we're seeking is researchers who find." The greater the confidence our leaders have in the quality of their researchers, the greater the latter's achievements will be.

Philippe Froguel
Geneticist at Imperial College London (ICL), a French expat.

I quit France for Britain in 2000, because I was frustrated by a system that refuses to recognize and reward success, and in which promotion can depend more on political skills than on scientific merit. Funding is not sufficiently tied to results, and evaluations fail to distinguish excellent from average research. Money is spread thinly among many institutions and fields, with insufficient focus on strengthening the best campuses and teams.

As chair of genomic medicine at ICL, I've discovered that a world-leading university can offer a unique range of opportunities to the most entrepreneurial academics. This is almost impossible to imagine in France, where universities and research institutions lack power, and are overly centralized and bureaucratic. Britain's academic world is unashamedly élitist, with constant evaluation of projects and outcomes. Institutions have long-term visions, and recruit a critical mass of the best scientists in order to achieve them. At ICL, individual success is rewarded in salaries and bonuses.

French universities should be given autonomy from the central government, which currently nominates chairs in medicine. Funding should be focused on just a handful of the best medical schools and university hospitals on the basis of a regular research assessment akin to the UK Research Assessment Exercise. And a single French Institute of Health should be created to end the inefficient competition among the many research players in the life sciences.

L. DALBAVIE

"The fact that the Grandes Ecoles divert the brightest young people away from science is catastrophic. We need to evolve the French system, taking such realities into account."
— Alain Trautmann

William Rutherford
Biophysicist at the Atomic Energy Commission's Photosystem II: Photochemistry, Water Oxidation and Regulation Institute in Saclay, near Paris; a British expat in France.

The French national sport is whingeing. They are very good at complaining about what's wrong with the research system. But they overstate the case. I did my PhD in London, and postdocs at the University of Illinois [at Urbana-Champaign] and the RIKEN Institute of Japan, and I think the French system is great in its own way.

The system is very different from elsewhere. There are peculiarities, such as the system of the Grandes Ecoles, and also much greater stability, in particular when starting out, because of the rolling grants funding system, and better job security. Coming here after my postdoc wanderings, I found it all very weird.

The system certainly has its weaknesses, and the French focus on these. But the truth is that there is much that is positive about the French system. Some things are better in the United States and Britain, but others are much better here. I really do believe that not having the huge pressure to publish before the next grant deadline allows you to do better science.

I also often see my US competitors, who are extremely good, go through patches in which they are less so. It's often to do with the loss of a grant or staff; they lose their expertise, and take time rebuilding it. Our lab, probably one of the best in the world for photosynthesis, doesn't have that problem. France is a great place to do what we do.

CELLECTIS

"The multiplication of publicly supported research organizations and structures has resulted in a failure to concentrate funds behind ambitious projects."
— André Choulika

There are downsides. Where should I start? France's legendary red tape is true bureaucracy for bureaucracy's sake. If you don't let it get to you, you can get your job done. Fixed salaries across the board are also a painful point, especially early in one's career.

Fixed salaries probably hit recruitment in sectors such as biotech, but are less of an issue in basic research. I could earn much more in Britain or the United States, but here I can do the science I want. Our lab also has no difficulty attracting foreign researchers. People love the lifestyle in France.

In reforming research, France must be careful not to throw the baby out with the bathwater. Much as I approve of certain aspects of the new ANR's competitive grant system, I certainly don't think that France should adopt an Anglo-Saxon system; the diversity of international systems is a richness that we should preserve. I would hate to see French research lose its unique character in an ill-advised attempt to copy the United States. ■

Declan Butler is a senior reporter for *Nature*, based in Paris.

See http://tinyurl.com/23fwj5 for comments from Bertrand Monthubert, president of Save Research, and Soumitra Dutta of business school INSEAD.

What do you think? Join in the debate on French science at http://blogs.nature.com/news/blog/2007/04/french_election.html

Let science speak for itself

Additional text for article doi:10.1038/446850a

Bertrand Monthubert
Mathematician at Paul Sabatier University in Toulouse and current president of 'Save Research'.

French researchers keep being told they should adopt an Anglo-Saxon system for research funding, yet if you look at Britain's share of world publications, its drop is, in fact, the same — 13% from 1999 to 2004. That said, we are not opposed to funding research on the basis of competitive projects. But we are opposed to the government's eagerness to use project funding for short-term strategic research, while at the same time cutting back on the research agencies that offer labs long-term stable funding.

INSU [the National Institute for Earth Sciences and Astronomy], for example, must plan expensive long-term ocean expeditions, and carry out its own peer review of missions. But current budgetary difficulties mean that it is having problems putting together expeditions for France's contribution to the International Polar Year.

You can't do this sort of research with 3-year project grants. The comparison of the French ANR to the US National Science Foundation [NSF] is inaccurate; the NSF often provides rolling funding for labs, much as the CNRS already does. France's centralized history allows us to carry out evaluations of laboratories at the national scale. The United States created its National Institutes of Health to obtain precisely that sort of overall vision, a common evaluation system. With the ANR, we risk giving labs large sums of money for 3 years, and then nothing more.

We support funding by project for new research areas in which France is lagging. We also support it as a way of bringing together the various research agencies. There are too many research agencies in France, and simplification is long overdue.

Increasingly, it is the science ministry that decides research orientations. They should leave it to the scientists to decide for themselves what the most important scientific opportunities are. Scientific progress cannot be made by decree.

In addition, instead of simplifying the structure of the research system, the government has multiplied new structures. We have hubs of competitiveness, thematic hubs and thematic networks. The list goes on. A single lab might be attached to four different hubs, all linked to others. Each of these structures has its own boards, meetings, statutes, funds, and requests for grant proposals. Just to get their job done, researchers have to juggle new layers of bureaucracy and all sorts of different calls for proposals.

French science does have serious difficulties, and there is worse to come, because the negative impact of recent government policies will take time to show.

Soumitra Dutta
Chair of business and technology at the European campus of INSEAD, a top business school in Fontainebleau, near Paris.

One should not underestimate France's strength in research and innovation. Innovation is much more complex than scores on this or that indicator. There are many factors that contribute to innovation, from infrastructure to human capacity, the ability to bring research to the market, and the complexity of doing business in a country.

At INSEAD this year we created a Global Innovation Index that combines more than 80 formal indicators on innovation, as well as informal surveys of the opinions of chief executives worldwide. France ranks fifth in the index, with the United States leading, followed by Germany, Britain and Japan.

France's strength lies in large technology-based companies in many sectors. It has been less successful in spawning new high-technology businesses, such as information technology. It also suffers from complexity in doing business, and inflexibility in its labour markets makes adapting to change difficult. Its lack of top research universities and the poor quality of tertiary education are weaknesses. Globalization means that success in innovation is now highly dependent on attracting top talent, and France isn't succeeding in this area.

France needs to realize that success in innovation also depends on factors outside France. Foreign talent is a critical factor, and no country can succeed without it, but talent is now globally mobile. If a country doesn't have the best universities it won't get the best students, and it won't have the best universities unless it can attract the best researchers.

An open attitude towards skilled immigration is also key, and this is going to be a massive problem for France's future. It ranks 61st in the world in terms of reducing obstacles to skilled foreign labour; Britain is 12th, a factor that largely explains its economic boom. Measures such as automatic work permits for skilled staff are a magnet for talent.

Immigration is such a politically charged issue in France that policies relating to it are a mess. France needs to be more open to skilled immigration, and needs to learn how to make foreigners feel at home — a key factor in the high-tech success of the United States. ■

Is French science in decline...

...or have its failings been greatly exaggerated? **Declan Butler** finds out.

Se regarder le nombril — navel gazing — is a national sport in France, and that gaze has turned to science. The sentiment that French research is in steep decline has become a recurrent theme of political discourse, newspaper editorials and TV talk shows.

France's research system has its problems — from unwieldy bureaucracies to dilapidated universities. As a result, many gaze across the channel, and the Atlantic, to the Anglo-Saxon system as a model for science and innovation.

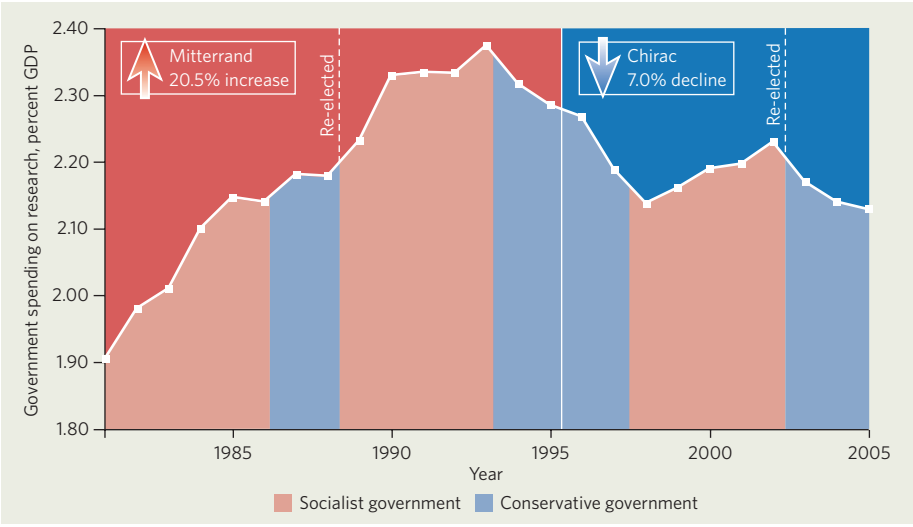
Key science indicators reveal a more complex picture, however. Despite deep funding cuts, basic research in France seems to be stable in terms of overall output of papers, but is losing ground in 'visibility' — the papers that have the greatest impact. The country does have a chronic weakness in private-sector research, though it is hardly alone in this malaise.

The challenge for French research in which politicians can make the biggest difference is the decline in government spending. Recent history reveals an 'accordion' effect in which right-leaning governments squeeze research funding, and socialist ones expand it (see graphic).

President François Mitterrand's election in 1981 marked the start of a golden decade for French research as spending grew rapidly under successive socialist governments. But the growth came to an abrupt end in 1993 when conservatives were elected under Prime Minister Edouard Balladur. And spending has fallen further since Jacques Chirac was elected as president in 1995, albeit punctuated by a rise under socialist prime minister Lionel Jospin from 1997 to 2002. But the net result of Chirac's 12-year presidency is that spending on science was lower as a proportion of GDP in 2005 — 2.13% — than it was in 1985.

This decline runs contrary to the goal of raising average science spending from 2% to 3% of GDP by 2010, which EU governments agreed in Lisbon in 2000. Some smaller European nations have already reached this target, although they remain the exceptions. Britain and Germany are the two European countries most directly comparable to France in research power, and the three together account for two-thirds of all research in the European Union. So how does France compare to them?

Germany overtook France as Europe's largest research spender in 1997, and its budget reached 2.52% of GDP in 2003. In the decade 1993–2003, UK spending fell by 8.7% to 1.88% — a similar drop to that seen in France.



SOURCE: OST, PARIS

But France has done better than Britain in private sector funding. Between 1993 and 2003, industry funding grew by 8% — close to Germany's 9%. By contrast, Britain has seen a 15% drop over this period — by 2003, to just 44% of its total science spending, well below the EU average of 54%. At 51%, France is very much an average European player and still falls short of Japan, where industry accounts for three-quarters of research funding, and the United States and Germany, where it accounts for two-thirds.

> **"French basic science is still holding its own."**
> — Ghislaine Filliatreau

How does all this spending translate into research outputs? France's share of the world's scientific papers had reached 5.4% by 1999, but by 2004 it had dropped to 4.7%, at which point it was displaced from fifth place in the world league by China. A comparable drop has occurred in most European countries, including Germany and Britain, and in the United States. "This negative rate is directly related to a positive growth rate in several countries, especially China," says Henk Moed, an expert in science indicators at Leiden University in the Netherlands.

The downward trends in research outputs is most marked in France's share of patents, particularly in the important US market. France's share of US patents dropped by 14% between 1999 and 2004 to 2.5%, whereas the EU's share as a whole fell by just 2% over this period to 17.1%. As with scientific publications, part of this drop is down to emerging economies in Asia. North America's share of US patents fell by 3% to 49.9% during this time; Asia's share rose by 6% to reach 30.1%, and South Korea overtook France and Britain with 2.6% of all US patents.

Of all the indicators, few can hurt French pride more than having so few institutions in the top rankings of international universities. But the poor performance of French universities is perhaps also exaggerated.

Under France's complex system, most labs belong to one or more research organizations and a university, but automated indicators such as Thomson's Essential Science Indicators count only the first affiliation in an author address, which means that French institutions are often not attributed correctly. To tackle this disparity, Thomson started a project in February with the Paris-based Observatoire des Sciences et des Techniques (OST) to get a truer picture of French science.

So the figures, although no cause for complacency, are not as bleak as some of the rhetoric might suggest. "We shouldn't overdramatize the supposed decline," says Ghislaine Filliatreau, director of the OST, "French basic science is still holding its own, despite poor funding, but risks being outspent and outperformed by other countries." ■

Declan Butler is a senior reporter for *Nature*, based in Paris.

Other riffs on cooperation are already showing how well a wiki could work

SIR — Barend Mons's Wiki for Professionals at www.wikiprofessional.info is among the first open collaborative databases to use the wiki format in biology, as your News story "Key biology databases go wiki" (*Nature* **445,** 691; 2007) points out.

However, other, non-wiki resources have already shown the feasibility of cooperative, online database construction. One such success story is GeneRIF (www.ncbi.nlm.nih.gov/projects/GeneRIF), which is like a miniature wiki where the author is restricted to a single short sentence. Currently, GeneRIF contains close to 200,000 entries, and each is attached to a particular gene at the Entrez database of the National Center for Biotechnology Information.

Being interested in gene–disease relationships, we assessed the coverage and specificity of GeneRIF and compared them to OMIM (Online Mendelian Inheritance in Man), a traditional source of gene-disease information. We found that GeneRIF already covers more than twice the number of diseases per gene and includes many more newly discovered mappings (www.basic.northwestern.edu/publications/generifdo). This seems to us to answer the scepticism that has been expressed about the expected community involvement in wiki collaborations.

John D. Osborne*, Simon Lin†, Warren A. Kibbe‡
*University of Alabama at Birmingham, 845 19th Street South, Bevill Building, Room 273C, Birmingham, Alabama 35294, USA
†Robert H. Lurie Comprehensive Cancer Center, Northwestern University, 676 North St Clair, Suite 1200, Chicago, Illinois 60611, USA
‡Center for Genetic Medicine, Robert H. Lurie Comprehensive Cancer Center and Feinberg School of Medicine, Northwestern University, 676 North St Clair, Suite 1200, Chicago, Illinois 60611, USA

Law and research could add up to profitable niche drugs

SIR — In response to your Editorial "A changing drug supply" (*Nature* **445,** 460; 2007), W. Ross Tracey points out in Correspondence that niche drugs may be just as expensive as blockbuster drugs to bring to the market, but a pharmaceutical company's revenue from them is likely to be only a fraction of what it can earn from selling blockbusters ("Niche drugs aren't a cheap alternative to blockbusters" *Nature* **445,** 818; 2007).

This is not necessarily so, in my opinion. Many drugs fail in clinical development, and some drugs have to be withdrawn after causing severe side effects in some patients. Once it becomes possible to determine what causes these patients to show adverse reactions, it may become feasible to pre-select a smaller patient population that will only respond positively and not suffer any adverse events. Although this strategy may not deliver blockbuster drugs, it could considerably reduce the present high attrition rate in drug development and thus make niche drugs economically viable.

Another point is that the US Food and Drug Administration uses financial incentives and an accelerated review process to support the development of drugs for 'orphan diseases' — conditions so uncommon that drug companies would otherwise have no incentive to seek cures for them (see the Orphan Drug Act, www.fda.gov/orphan/oda.htm).

Burkhard Haefner
Johnson & Johnson Pharma Research and Development, Turnhoutseweg 30, Box 6423, 2340 Beerse, Belgium

Why do so few women speak at science meetings?

SIR — Mary Ann Holmes and Suzanne O'Connell comment on the lack of women in the academic ranks in your Recruiters article "Leaks in the pipeline" (*Nature* **446,** 346; 2007). In the same issue, advertisements for two Nature conferences illustrate part of the problem — the poor representation of women speakers at scientific meetings.

The Nature conference "Oncogenes and human cancer: the next 25 years" features 36 speakers, of whom four are women. The "Days of molecular medicine: emerging technologies and cancer biology" conference, co-sponsored by *Nature Medicine*, features 19 speakers, of whom two are women. There are many accomplished women scientists in the areas covered by these meetings. There is no obvious reason why the number of women speakers should be so low.

The representation of women speakers at many meetings remains dismally poor and thus may contribute to the lack of success of women in academia.

However, this is a problem that could be easily remedied, if more attention were paid by organizers and the agencies that provide funding for meetings to the issue of whether qualified female speakers have been missed.

Pamela A. Silver
Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

Who will start the 3Rs ball rolling for animal welfare?

SIR — Hanno Würbel, in his Correspondence "Publications should include an animal-welfare section", suggests an effective and powerful way in which journals, by including a dedicated category for the 3Rs — replace, refine, reduce — in the methodology section, could benefit both scientific research and animal welfare (*Nature* **446,** 257; 2007, and see http://blogs.nature.com/nautilus/2007/03/proposal_for_journals_to_inclu.html) .

Ever since this idea was recommended by the Nuffield Council on Bioethics, I have been probing scientists for their response to it. Almost without exception, those I have approached have recognized its potential and support the idea verbally. However, they have consistently been reticent about providing any written endorsement.

The first, but, in my opinion, least likely, reason for this behaviour could be apathy, because this sensible, moderate and pragmatic proposal does not arouse the same emotions as the 'animal research' debate itself.

Second, the polarized nature of this topic may foster the fear that publicly supporting any measure to improve animal welfare will be perceived as a defection by other scientists.

Third, there may be an unwillingness to admit what some would see as a weakening of stance — the thin end of a wedge that threatens to phase out animal research altogether. Although my personal desire is that this would indeed be the outcome, the evidence speaks otherwise. Despite the best efforts of animal-welfare advocates since the 3Rs concept was first introduced, 'replacements' have become established only when they are of scientific benefit — so it seems unreasonable that efforts to encourage them should be viewed as a threat to progress.

Finally, there is resistance to going against the grain. If a single high-impact journal were to take a unilateral decision to implement this proposal and embrace the 3Rs as integral and essential elements of good experimental design, I believe that other editors and scientists everywhere would be happy to follow.

Victoria Buck
Division of Stem Cell Biology and Developmental Genetics, MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

Comments are welcome at Nautilus, the blog for authors, at the URL above.

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 500 words, and ideally shorter.

Other riffs on cooperation are already showing how well a wiki could work

SIR — Barend Mons's Wiki for Professionals at www.wikiprofessional.info is among the first open collaborative databases to use the wiki format in biology, as your News story "Key biology databases go wiki" (*Nature* **445,** 691; 2007) points out.

However, other, non-wiki resources have already shown the feasibility of cooperative, online database construction. One such success story is GeneRIF (www.ncbi.nlm.nih.gov/projects/GeneRIF), which is like a miniature wiki where the author is restricted to a single short sentence. Currently, GeneRIF contains close to 200,000 entries, and each is attached to a particular gene at the Entrez database of the National Center for Biotechnology Information.

Being interested in gene–disease relationships, we assessed the coverage and specificity of GeneRIF and compared them to OMIM (Online Mendelian Inheritance in Man), a traditional source of gene-disease information. We found that GeneRIF already covers more than twice the number of diseases per gene and includes many more newly discovered mappings (www.basic.northwestern.edu/publications/generifdo). This seems to us to answer the scepticism that has been expressed about the expected community involvement in wiki collaborations.

John D. Osborne*, Simon Lin†, Warren A. Kibbe‡
*University of Alabama at Birmingham, 845 19th Street South, Bevill Building, Room 273C, Birmingham, Alabama 35294, USA
†Robert H. Lurie Comprehensive Cancer Center, Northwestern University, 676 North St Clair, Suite 1200, Chicago, Illinois 60611, USA
‡Center for Genetic Medicine, Robert H. Lurie Comprehensive Cancer Center and Feinberg School of Medicine, Northwestern University, 676 North St Clair, Suite 1200, Chicago, Illinois 60611, USA

Law and research could add up to profitable niche drugs

SIR — In response to your Editorial "A changing drug supply" (*Nature* **445,** 460; 2007), W. Ross Tracey points out in Correspondence that niche drugs may be just as expensive as blockbuster drugs to bring to the market, but a pharmaceutical company's revenue from them is likely to be only a fraction of what it can earn from selling blockbusters ("Niche drugs aren't a cheap alternative to blockbusters" *Nature* **445,** 818; 2007).

This is not necessarily so, in my opinion. Many drugs fail in clinical development, and some drugs have to be withdrawn after causing severe side effects in some patients. Once it becomes possible to determine what causes these patients to show adverse reactions, it may become feasible to pre-select a smaller patient population that will only respond positively and not suffer any adverse events. Although this strategy may not deliver blockbuster drugs, it could considerably reduce the present high attrition rate in drug development and thus make niche drugs economically viable.

Another point is that the US Food and Drug Administration uses financial incentives and an accelerated review process to support the development of drugs for 'orphan diseases' — conditions so uncommon that drug companies would otherwise have no incentive to seek cures for them (see the Orphan Drug Act, www.fda.gov/orphan/oda.htm).

Burkhard Haefner
Johnson & Johnson Pharma Research and Development, Turnhoutseweg 30, Box 6423, 2340 Beerse, Belgium

Why do so few women speak at science meetings?

SIR — Mary Ann Holmes and Suzanne O'Connell comment on the lack of women in the academic ranks in your Recruiters article "Leaks in the pipeline" (*Nature* **446,** 346; 2007). In the same issue, advertisements for two Nature conferences illustrate part of the problem — the poor representation of women speakers at scientific meetings.

The Nature conference "Oncogenes and human cancer: the next 25 years" features 36 speakers, of whom four are women. The "Days of molecular medicine: emerging technologies and cancer biology" conference, co-sponsored by *Nature Medicine*, features 19 speakers, of whom two are women. There are many accomplished women scientists in the areas covered by these meetings. There is no obvious reason why the number of women speakers should be so low.

The representation of women speakers at many meetings remains dismally poor and thus may contribute to the lack of success of women in academia.

However, this is a problem that could be easily remedied, if more attention were paid by organizers and the agencies that provide funding for meetings to the issue of whether qualified female speakers have been missed.

Pamela A. Silver
Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

Who will start the 3Rs ball rolling for animal welfare?

SIR — Hanno Würbel, in his Correspondence "Publications should include an animal-welfare section", suggests an effective and powerful way in which journals, by including a dedicated category for the 3Rs — replace, refine, reduce — in the methodology section, could benefit both scientific research and animal welfare (*Nature* **446,** 257; 2007, and see http://blogs.nature.com/nautilus/2007/03/proposal_for_journals_to_inclu.html) .

Ever since this idea was recommended by the Nuffield Council on Bioethics, I have been probing scientists for their response to it. Almost without exception, those I have approached have recognized its potential and support the idea verbally. However, they have consistently been reticent about providing any written endorsement.

The first, but, in my opinion, least likely, reason for this behaviour could be apathy, because this sensible, moderate and pragmatic proposal does not arouse the same emotions as the 'animal research' debate itself.

Second, the polarized nature of this topic may foster the fear that publicly supporting any measure to improve animal welfare will be perceived as a defection by other scientists.

Third, there may be an unwillingness to admit what some would see as a weakening of stance — the thin end of a wedge that threatens to phase out animal research altogether. Although my personal desire is that this would indeed be the outcome, the evidence speaks otherwise. Despite the best efforts of animal-welfare advocates since the 3Rs concept was first introduced, 'replacements' have become established only when they are of scientific benefit — so it seems unreasonable that efforts to encourage them should be viewed as a threat to progress.

Finally, there is resistance to going against the grain. If a single high-impact journal were to take a unilateral decision to implement this proposal and embrace the 3Rs as integral and essential elements of good experimental design, I believe that other editors and scientists everywhere would be happy to follow.

Victoria Buck
Division of Stem Cell Biology and Developmental Genetics, MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

Comments are welcome at Nautilus, the blog for authors, at the URL above.

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 500 words, and ideally shorter.

Other riffs on cooperation are already showing how well a wiki could work

SIR — Barend Mons's Wiki for Professionals at www.wikiprofessional.info is among the first open collaborative databases to use the wiki format in biology, as your News story "Key biology databases go wiki" (*Nature* **445,** 691; 2007) points out.

However, other, non-wiki resources have already shown the feasibility of cooperative, online database construction. One such success story is GeneRIF (www.ncbi.nlm.nih.gov/projects/GeneRIF), which is like a miniature wiki where the author is restricted to a single short sentence. Currently, GeneRIF contains close to 200,000 entries, and each is attached to a particular gene at the Entrez database of the National Center for Biotechnology Information.

Being interested in gene–disease relationships, we assessed the coverage and specificity of GeneRIF and compared them to OMIM (Online Mendelian Inheritance in Man), a traditional source of gene-disease information. We found that GeneRIF already covers more than twice the number of diseases per gene and includes many more newly discovered mappings (www.basic.northwestern.edu/publications/generifdo). This seems to us to answer the scepticism that has been expressed about the expected community involvement in wiki collaborations.

John D. Osborne*, Simon Lin†, Warren A. Kibbe‡
*University of Alabama at Birmingham, 845 19th Street South, Bevill Building, Room 273C, Birmingham, Alabama 35294, USA
†Robert H. Lurie Comprehensive Cancer Center, Northwestern University, 676 North St Clair, Suite 1200, Chicago, Illinois 60611, USA
‡Center for Genetic Medicine, Robert H. Lurie Comprehensive Cancer Center and Feinberg School of Medicine, Northwestern University, 676 North St Clair, Suite 1200, Chicago, Illinois 60611, USA

Law and research could add up to profitable niche drugs

SIR — In response to your Editorial "A changing drug supply" (*Nature* **445,** 460; 2007), W. Ross Tracey points out in Correspondence that niche drugs may be just as expensive as blockbuster drugs to bring to the market, but a pharmaceutical company's revenue from them is likely to be only a fraction of what it can earn from selling blockbusters ("Niche drugs aren't a cheap alternative to blockbusters" *Nature* **445,** 818; 2007).

This is not necessarily so, in my opinion. Many drugs fail in clinical development, and some drugs have to be withdrawn after causing severe side effects in some patients. Once it becomes possible to determine what causes these patients to show adverse reactions, it may become feasible to pre-select a smaller patient population that will only respond positively and not suffer any adverse events. Although this strategy may not deliver blockbuster drugs, it could considerably reduce the present high attrition rate in drug development and thus make niche drugs economically viable.

Another point is that the US Food and Drug Administration uses financial incentives and an accelerated review process to support the development of drugs for 'orphan diseases' — conditions so uncommon that drug companies would otherwise have no incentive to seek cures for them (see the Orphan Drug Act, www.fda.gov/orphan/oda.htm).

Burkhard Haefner
Johnson & Johnson Pharma Research and Development, Turnhoutseweg 30, Box 6423, 2340 Beerse, Belgium

Why do so few women speak at science meetings?

SIR — Mary Ann Holmes and Suzanne O'Connell comment on the lack of women in the academic ranks in your Recruiters article "Leaks in the pipeline" (*Nature* **446,** 346; 2007). In the same issue, advertisements for two Nature conferences illustrate part of the problem — the poor representation of women speakers at scientific meetings.

The Nature conference "Oncogenes and human cancer: the next 25 years" features 36 speakers, of whom four are women. The "Days of molecular medicine: emerging technologies and cancer biology" conference, co-sponsored by *Nature Medicine*, features 19 speakers, of whom two are women. There are many accomplished women scientists in the areas covered by these meetings. There is no obvious reason why the number of women speakers should be so low.

The representation of women speakers at many meetings remains dismally poor and thus may contribute to the lack of success of women in academia.

However, this is a problem that could be easily remedied, if more attention were paid by organizers and the agencies that provide funding for meetings to the issue of whether qualified female speakers have been missed.

Pamela A. Silver
Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

Who will start the 3Rs ball rolling for animal welfare?

SIR — Hanno Würbel, in his Correspondence "Publications should include an animal-welfare section", suggests an effective and powerful way in which journals, by including a dedicated category for the 3Rs — replace, refine, reduce — in the methodology section, could benefit both scientific research and animal welfare (*Nature* **446,** 257; 2007, and see http://blogs.nature.com/nautilus/2007/03/proposal_for_journals_to_inclu.html) .

Ever since this idea was recommended by the Nuffield Council on Bioethics, I have been probing scientists for their response to it. Almost without exception, those I have approached have recognized its potential and support the idea verbally. However, they have consistently been reticent about providing any written endorsement.

The first, but, in my opinion, least likely, reason for this behaviour could be apathy, because this sensible, moderate and pragmatic proposal does not arouse the same emotions as the 'animal research' debate itself.

Second, the polarized nature of this topic may foster the fear that publicly supporting any measure to improve animal welfare will be perceived as a defection by other scientists.

Third, there may be an unwillingness to admit what some would see as a weakening of stance — the thin end of a wedge that threatens to phase out animal research altogether. Although my personal desire is that this would indeed be the outcome, the evidence speaks otherwise. Despite the best efforts of animal-welfare advocates since the 3Rs concept was first introduced, 'replacements' have become established only when they are of scientific benefit — so it seems unreasonable that efforts to encourage them should be viewed as a threat to progress.

Finally, there is resistance to going against the grain. If a single high-impact journal were to take a unilateral decision to implement this proposal and embrace the 3Rs as integral and essential elements of good experimental design, I believe that other editors and scientists everywhere would be happy to follow.

Victoria Buck
Division of Stem Cell Biology and Developmental Genetics, MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

Comments are welcome at Nautilus, the blog for authors, at the URL above.

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 500 words, and ideally shorter.

Other riffs on cooperation are already showing how well a wiki could work

SIR — Barend Mons's Wiki for Professionals at www.wikiprofessional.info is among the first open collaborative databases to use the wiki format in biology, as your News story "Key biology databases go wiki" (*Nature* **445,** 691; 2007) points out.

However, other, non-wiki resources have already shown the feasibility of cooperative, online database construction. One such success story is GeneRIF (www.ncbi.nlm.nih.gov/projects/GeneRIF), which is like a miniature wiki where the author is restricted to a single short sentence. Currently, GeneRIF contains close to 200,000 entries, and each is attached to a particular gene at the Entrez database of the National Center for Biotechnology Information.

Being interested in gene–disease relationships, we assessed the coverage and specificity of GeneRIF and compared them to OMIM (Online Mendelian Inheritance in Man), a traditional source of gene-disease information. We found that GeneRIF already covers more than twice the number of diseases per gene and includes many more newly discovered mappings (www.basic.northwestern.edu/publications/generifdo). This seems to us to answer the scepticism that has been expressed about the expected community involvement in wiki collaborations.

John D. Osborne*, Simon Lin†, Warren A. Kibbe‡
*University of Alabama at Birmingham, 845 19th Street South, Bevill Building, Room 273C, Birmingham, Alabama 35294, USA
†Robert H. Lurie Comprehensive Cancer Center, Northwestern University, 676 North St Clair, Suite 1200, Chicago, Illinois 60611, USA
‡Center for Genetic Medicine, Robert H. Lurie Comprehensive Cancer Center and Feinberg School of Medicine, Northwestern University, 676 North St Clair, Suite 1200, Chicago, Illinois 60611, USA

Law and research could add up to profitable niche drugs

SIR — In response to your Editorial "A changing drug supply" (*Nature* **445,** 460; 2007), W. Ross Tracey points out in Correspondence that niche drugs may be just as expensive as blockbuster drugs to bring to the market, but a pharmaceutical company's revenue from them is likely to be only a fraction of what it can earn from selling blockbusters ("Niche drugs aren't a cheap alternative to blockbusters" *Nature* **445,** 818; 2007).

This is not necessarily so, in my opinion. Many drugs fail in clinical development, and some drugs have to be withdrawn after causing severe side effects in some patients. Once it becomes possible to determine what causes these patients to show adverse reactions, it may become feasible to pre-select a smaller patient population that will only respond positively and not suffer any adverse events. Although this strategy may not deliver blockbuster drugs, it could considerably reduce the present high attrition rate in drug development and thus make niche drugs economically viable.

Another point is that the US Food and Drug Administration uses financial incentives and an accelerated review process to support the development of drugs for 'orphan diseases' — conditions so uncommon that drug companies would otherwise have no incentive to seek cures for them (see the Orphan Drug Act, www.fda.gov/orphan/oda.htm).

Burkhard Haefner
Johnson & Johnson Pharma Research and Development, Turnhoutseweg 30, Box 6423, 2340 Beerse, Belgium

Why do so few women speak at science meetings?

SIR — Mary Ann Holmes and Suzanne O'Connell comment on the lack of women in the academic ranks in your Recruiters article "Leaks in the pipeline" (*Nature* **446,** 346; 2007). In the same issue, advertisements for two Nature conferences illustrate part of the problem — the poor representation of women speakers at scientific meetings.

The Nature conference "Oncogenes and human cancer: the next 25 years" features 36 speakers, of whom four are women. The "Days of molecular medicine: emerging technologies and cancer biology" conference, co-sponsored by *Nature Medicine*, features 19 speakers, of whom two are women. There are many accomplished women scientists in the areas covered by these meetings. There is no obvious reason why the number of women speakers should be so low.

The representation of women speakers at many meetings remains dismally poor and thus may contribute to the lack of success of women in academia.

However, this is a problem that could be easily remedied, if more attention were paid by organizers and the agencies that provide funding for meetings to the issue of whether qualified female speakers have been missed.

Pamela A. Silver
Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

Who will start the 3Rs ball rolling for animal welfare?

SIR — Hanno Würbel, in his Correspondence "Publications should include an animal-welfare section", suggests an effective and powerful way in which journals, by including a dedicated category for the 3Rs — replace, refine, reduce — in the methodology section, could benefit both scientific research and animal welfare (*Nature* **446,** 257; 2007, and see http://blogs.nature.com/nautilus/2007/03/proposal_for_journals_to_inclu.html) .

Ever since this idea was recommended by the Nuffield Council on Bioethics, I have been probing scientists for their response to it. Almost without exception, those I have approached have recognized its potential and support the idea verbally. However, they have consistently been reticent about providing any written endorsement.

The first, but, in my opinion, least likely, reason for this behaviour could be apathy, because this sensible, moderate and pragmatic proposal does not arouse the same emotions as the 'animal research' debate itself.

Second, the polarized nature of this topic may foster the fear that publicly supporting any measure to improve animal welfare will be perceived as a defection by other scientists.

Third, there may be an unwillingness to admit what some would see as a weakening of stance — the thin end of a wedge that threatens to phase out animal research altogether. Although my personal desire is that this would indeed be the outcome, the evidence speaks otherwise. Despite the best efforts of animal-welfare advocates since the 3Rs concept was first introduced, 'replacements' have become established only when they are of scientific benefit — so it seems unreasonable that efforts to encourage them should be viewed as a threat to progress.

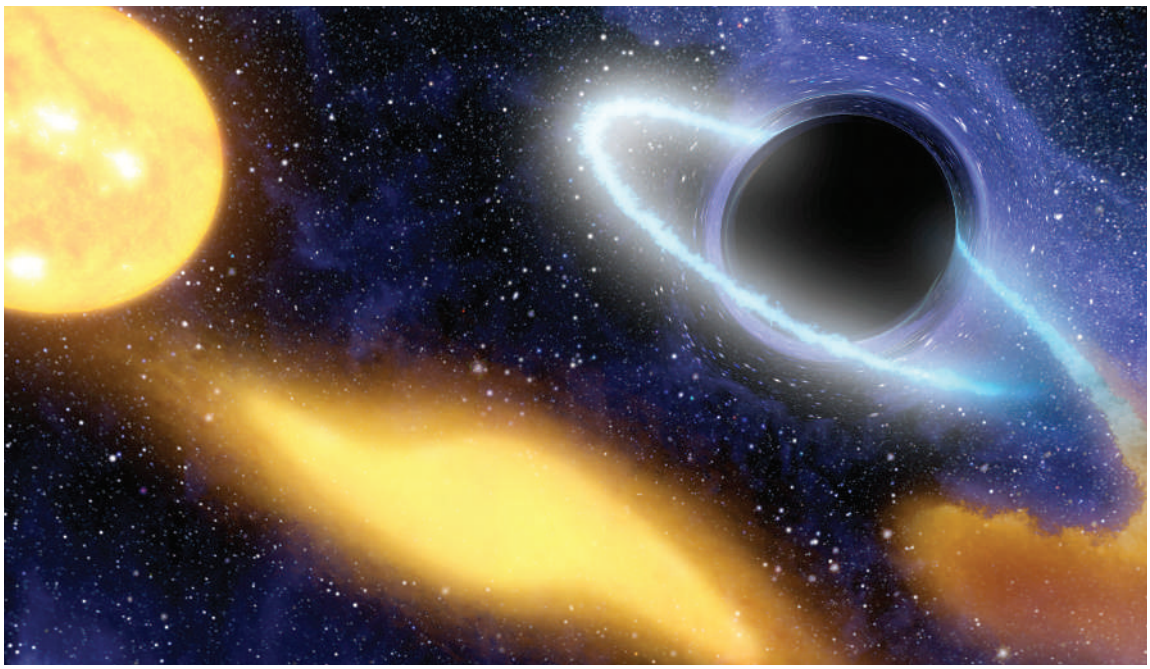
Finally, there is resistance to going against the grain. If a single high-impact journal were to take a unilateral decision to implement this proposal and embrace the 3Rs as integral and essential elements of good experimental design, I believe that other editors and scientists everywhere would be happy to follow.

Victoria Buck
Division of Stem Cell Biology and Developmental Genetics, MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

Comments are welcome at Nautilus, the blog for authors, at the URL above.

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 500 words, and ideally shorter.

BOOKS & ARTS



Einstein's general theory of relativity sparked research that led to the understanding of black holes.

NASA-JPL

Relative confusion

The implications of Einstein's special and general theories of relativity were not immediately apparent.

The Curious History of Relativity: How Einstein's Theory of Gravity was Lost and Found Again
by Jean Eisenstaedt
Princeton University Press: 2006. 384 pp. $29.95, £18.95

William Unruh

Einstein created two surprising theories between 1905 and 1915. His special theory of relativity in 1905 altered our understanding of space and time such that, rather than being viewed as aspects of the physical world that could be used to describe and explain other features, both came to depend on the observer. By 1915, his general theory of relativity meant that space and time were now even more fluid, also depending on the state of matter in the surrounding Universe. Gravity was now seen as an aspect of time and space.

The Curious History of Relativity by Jean Eisenstaedt, originally published in French as *Einstein et la Relativité Générale*, traces Einstein's development of his special and general theories of relativity. It also describes the confused state of the experiments in the following years that were designed to test the general theory. The final third of the book describes and explains one of the consequences of the theory: the idea of black holes. As a historian of physics who has studied the development of general relativity, Eisenstaedt presents a wealth of intellectual, social and philosophical material regarding the theory and the reactions it engendered in the community.

I found many parts of this book absorbing, but some were not without irritation. Let me start with the positives. Eisenstaedt's mastery of Einstein's intellectual and personal development between 1905 and 1920, a period that spanned his greatest creativity and contribution, is evident, and he tells the stories well. It is fascinating to read his descriptions of Einstein's struggles with the concepts, his tenacious grip on the principles that helped him escape from numerous dead ends, and his temporary falling out with David Hilbert over what he regarded (justifiably) as an attempt by Hilbert at intellectual theft. The author's interesting digressions on the nature of scientific research and on the way scientific theories are invented are appropriate, even if I did not always agree with his position. His emphasis that both special and general relativity are theories of what is invariant, rather than what is relative, is spot on.

However, parts of the book annoyed me. Before describing the problems, I must declare a conflict: Eisenstaedt implies that relativists, and I consider myself one, are confused, incompetent and studying the subject just because they are paid to. Nevertheless, I will outline aspects of this book that bothered me. First, Eisenstaedt masterfully recounts the conventional story about the puzzles concerning the behaviour of light that arose in the nineteenth century. His treatment of these is fascinating in its historical detail, but he implies that half a century of trouble could have been avoided if only the physicists of the time had realized that velocities should not be added together. In contrast, even with the special theory of relativity, relative velocities as seen by any one observer should be added. The key to special relativity is that observers disagree on what those velocities are because they differ on what distances and time mean. This point is subtle, and escaped even Henri Poincaré and Hendrik Lorentz shortly before 1905.

Eisenstaedt fails to convey an intuitive understanding of general relativity. He resorts to a statement that gravity is the curvature of space-time — an explanation guaranteed to shut off any reader's imagination — and his attempts to describe tensor calculus do not help. The primary image he uses is that of the deflections experienced by a ball rolling across a rumpled landscape. But the primary cause of those deflections is the effect of Earth's gravity, so his description is circular as it uses gravity to explain gravity. He describes the Einstein

effect, whereby clocks run at different rates according to their position in a gravitational field, but gravity does not cause this effect. Rather, general relativity says that this inequable flow of time from place to place is gravity, in most of our normal experiences with it.

My biggest gripe about the book, however, is the conceit in the title that the theory was almost lost in the 1930s and 1940s, and was only found again in the early 1960s with the discovery of black holes. It is true that during the Second World War everyone's attention was concentrated on more practical areas of science, but it was precisely during that period that general relativity completely revised our image of the Universe. In conflict with even Einstein's prejudices, the Universe as a whole was found to be dynamic. It was also during this period that Subrahmanyan Chandrasekhar, and later Robert Oppenheimer and his students, used the theory to show that large stars at the end of their lives must become what we now know as black holes. Immediately after the war, groups across the world started studying the physical and mathematical consequences of the theory. Black holes did not cause that revival; understanding them was its result. Eisenstaedt's sensationalization is unnecessary in a book about what is one of the most sensational intellectual stories of the twentieth century.

Despite these grumbles, I would recommend the book for its account of those ten incredible years and the impact that was generated. ■

William Unruh is in the Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada.

A hunger for company

Feast: Why Humans Share Food
by Martin Jones
Oxford University Press: 2007. 364 pp.
$35, £20

Gary Paul Nabhan

Engaging storytelling is not the forte of many technical scholars. So when an intelligent book comes along that is also truly charming, it deserves celebration. *Feast* by Martin Jones, a bio-archaeologist at the University of Cambridge, will delight most anthropologists and evolutionary biologists, as well as broadly educated laypersons interested in the evolution of diet and the social organization of eating.

The book is a pertinent example of what can be gained by 'consilience' among the natural sciences, social sciences and humanities. The term 'consilience' took on an extra layer of meaning when Edward O. Wilson attempted what might be perceived as a hostile take-over of all such disciplines by subsuming them under the unifying principles of evolutionary biology. In contrast, Jones takes a more balanced approach, setting up a productive tension between the evolutionary theories of Charles Darwin and Marvin Harris on the one hand, and the social and metaphorical insights of Mary Douglas and Claude Lévi-Strauss on the other. Include the biomolecular toolkit that Jones has used to look at remnant foodstuffs found on ancient grinding stones, clay pots and teeth, and you have an entirely fresh, integrated view of the social dynamics of food harvesting, food preparation and eating.

One of the most satisfying aspects of this treatise on the evolution of socialized food preparation and consumption is the manner in which Jones demonstrates the utility of new instruments, techniques and methodologies for investigating fragmentary archaeological remains associated with sites of human food procurement. Rather than assuming that these technologies render other, earlier approaches to human dietary evolution obsolete, Jones realizes that they build on the work of earlier investigators. Tool-driven science can go astray if it is not grounded in the well-articulated theories that it not only tests, but also refines. Jones therefore offers a narrative that is simultaneously humble and excited by new opportunities to shed light on why humans eat in the ways they do.

The journey makes stopovers at ancient archaeological sites; at the African field camps where Jane Goodall and other ethologists observed the eating behaviours of different primates; and even at the Fresh Kills landfill site near New York City, where garbologist William Rathje analyses contemporary human behaviour by sifting through masses of kitchen waste. Some of these stopovers have also been visited by other scholars on their quest to understand the social behaviour of human food consumers, and Jones uses each of their parables to illustrate his cogent points.

Jones even records field notes on his own ritualized behaviour and dress at a banquet at the University of Cambridge in the framework of their social acceptability in this formal setting. By juxtaposing this with primate behaviour at Goodall's camp, he offers a compelling answer to the question of whether humans are really that different from other primates in the social structure of our food sharing. The short answer is no, we are not, but the richer answer is that it is a matter of degree, and Jones guides us through these nuances.

By concerning himself with both evolutionary theory and the social constructions that give meaning to human eating behaviour, Jones arrives at a robust, integrative theory of why we share food the way we do. Testing this theory is likely to keep interdisciplinary scholars engaged for several decades to come. In a model of accessible scientific writing, Jones' captivating narrative is based on cutting-edge technology and on his personal indebtedness to early pioneers in this field. ■

Gary Paul Nabhan directs the Renewing America's Food Traditions consortium at the Center for Sustainable Environments, Northern Arizona University, Flagstaff, Arizona 86011, USA.

L. PSIHOYOS/SCIENCE FACTION

Your lab's a tip: garbologist William Rathje finds clues to human behaviour in piles of rubbish.

effect, whereby clocks run at different rates according to their position in a gravitational field, but gravity does not cause this effect. Rather, general relativity says that this inequable flow of time from place to place is gravity, in most of our normal experiences with it.

My biggest gripe about the book, however, is the conceit in the title that the theory was almost lost in the 1930s and 1940s, and was only found again in the early 1960s with the discovery of black holes. It is true that during the Second World War everyone's attention was concentrated on more practical areas of science, but it was precisely during that period that general relativity completely revised our image of the Universe. In conflict with even Einstein's prejudices, the Universe as a whole was found to be dynamic. It was also during this period that Subrahmanyan Chandrasekhar, and later Robert Oppenheimer and his students, used the theory to show that large stars at the end of their lives must become what we now know as black holes. Immediately after the war, groups across the world started studying the physical and mathematical consequences of the theory. Black holes did not cause that revival; understanding them was its result. Eisenstaedt's sensationalization is unnecessary in a book about what is one of the most sensational intellectual stories of the twentieth century.

Despite these grumbles, I would recommend the book for its account of those ten incredible years and the impact that was generated. ■

William Unruh is in the Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada.

A hunger for company

Feast: Why Humans Share Food
by Martin Jones
Oxford University Press: 2007. 364 pp.
$35, £20

Gary Paul Nabhan

Engaging storytelling is not the forte of many technical scholars. So when an intelligent book comes along that is also truly charming, it deserves celebration. *Feast* by Martin Jones, a bio-archaeologist at the University of Cambridge, will delight most anthropologists and evolutionary biologists, as well as broadly educated laypersons interested in the evolution of diet and the social organization of eating.

The book is a pertinent example of what can be gained by 'consilience' among the natural sciences, social sciences and humanities. The term 'consilience' took on an extra layer of meaning when Edward O. Wilson attempted what might be perceived as a hostile take-over of all such disciplines by subsuming them under the unifying principles of evolutionary biology. In contrast, Jones takes a more balanced approach, setting up a productive tension between the evolutionary theories of Charles Darwin and Marvin Harris on the one hand, and the social and metaphorical insights of Mary Douglas and Claude Lévi-Strauss on the other. Include the biomolecular toolkit that Jones has used to look at remnant foodstuffs found on ancient grinding stones, clay pots and teeth, and you have an entirely fresh, integrated view of the social dynamics of food harvesting, food preparation and eating.

One of the most satisfying aspects of this treatise on the evolution of socialized food preparation and consumption is the manner in which Jones demonstrates the utility of new instruments, techniques and methodologies for investigating fragmentary archaeological remains associated with sites of human food procurement. Rather than assuming that these technologies render other, earlier approaches to human dietary evolution obsolete, Jones realizes that they build on the work of earlier investigators. Tool-driven science can go astray if it is not grounded in the well-articulated theories that it not only tests, but also refines. Jones therefore offers a narrative that is simultaneously humble and excited by new opportunities to shed light on why humans eat in the ways they do.

The journey makes stopovers at ancient archaeological sites; at the African field camps where Jane Goodall and other ethologists observed the eating behaviours of different primates; and even at the Fresh Kills landfill site near New York City, where garbologist William Rathje analyses contemporary human behaviour by sifting through masses of kitchen waste. Some of these stopovers have also been visited by other scholars on their quest to understand the social behaviour of human food consumers, and Jones uses each of their parables to illustrate his cogent points.

Jones even records field notes on his own ritualized behaviour and dress at a banquet at the University of Cambridge in the framework of their social acceptability in this formal setting. By juxtaposing this with primate behaviour at Goodall's camp, he offers a compelling answer to the question of whether humans are really that different from other primates in the social structure of our food sharing. The short answer is no, we are not, but the richer answer is that it is a matter of degree, and Jones guides us through these nuances.

By concerning himself with both evolutionary theory and the social constructions that give meaning to human eating behaviour, Jones arrives at a robust, integrative theory of why we share food the way we do. Testing this theory is likely to keep interdisciplinary scholars engaged for several decades to come. In a model of accessible scientific writing, Jones' captivating narrative is based on cutting-edge technology and on his personal indebtedness to early pioneers in this field. ■

Gary Paul Nabhan directs the Renewing America's Food Traditions consortium at the Center for Sustainable Environments, Northern Arizona University, Flagstaff, Arizona 86011, USA.

L. PSIHOYOS/SCIENCE FACTION

Your lab's a tip: garbologist William Rathje finds clues to human behaviour in piles of rubbish.

THEATRE

Here's looking at you

C. ROSEGG

The ninth annual season of the First Light Festival is under way in New York, offering new dramatic works exploring science and technology, commissioned by the Ensemble Studio Theatre and the Alfred P. Sloan Foundation. The centrepiece is the premiere of *Serendib*, a play by David Zellnik, inspired by two months the author spent with a team of primatologists studying toque macaques in Sri Lanka. The play focuses on scientists George Fischke and Anna Sunilagatte, who try to test the hypothesis that occasional violent takeovers of one troop of macaques by an outside alpha male adversely affects the troop's happiness. But money is short, so the scientists turn to two vaguely disreputable documentary film-makers for some much-needed publicity. The film-makers bring along Dmitri Ramsov, a geneticist who has no use for Fischke's woolly theories about the soulfulness of macaques. The stage is thus set for two mirror-image plays, with Fischke and Ramsov vying for supremacy and the affections of Sunilagatte, and the macaques under observation — puppets controlled by the actors — acting out a parallel drama. The film-makers' search for good television provides comic relief, but the play persists with its themes: the roles of observation and experiment in science, the distorting lens of media coverage, and the debate over the relative importance of heredity and environment in behaviour. It reaches a moving climax as Fischke and Ramsov deliver eloquent tributes to two of the macaques, their words illustrating the differences between humans and our primate cousins, and the powerful, if elusive, similarities. **Alan Packer**

http://ensemblestudiotheatre.org/index3.html

No firm promises

Science Business: The Promise, the Reality, and the Future of Biotech
by Gary P. Pisano
Harvard Business School Press: 2006.
256 pp. $29.95

Keith Redpath

I have an aversion to management books, and it is based on the way they are written. Why use a sentence when a chapter will do? Why use standard English when you can invent meaningless phrases? Gary Pisano suffers from this predilection in his book *Science Business*. He takes more than 70 pages to tell the reader that drug discovery is complex and risky, involving several disciplines and lots of money, and that biotechnology has simply increased the complexity. A further 30 pages are required to conclude that many commercialization strategies are being followed, but that it is too early to know which are right.

What Pisano does, however, which few have dared to do in print before, is the biotechnology equivalent of the boy who shouted out that the Emperor had no clothes. He poses the question: "can science be a business?" He concludes, on the basis of his research and the experience of biotechnology to date, that the answer is no, but that the industry still has time to exonerate itself, if only it can change.

Modern biotechnology was depicted at its inception as a huge danger to mankind. Genetic manipulation was either carried out behind greater security than was required for the reprocessing of nuclear waste, or it was banned altogether. We were told that mad scientists would release modified microbes that would wipe out mankind. The hysteria parallels that accompanying stem cells and nanotechnology some 30 years later and perhaps demonstrates one of Pisano's theses: that the industry fails to learn from its mistakes.

Biotechnology was soon repositioned as the saviour of mankind. No disease was immune to attack from the magic bullet of biotechnology. Shares soared. Investors cashed in and got rich. Many more companies formed. Universities saw ways of making money from research.

Despite that, Pisano contends that the industry has failed to deliver financially. He may be right, but his analysis is flawed. Many small start-up companies with promising drug candidates or technologies are acquired by larger companies before making a profit, delivering good returns for founders and investors. In deciding whether biotechnology has delivered financially, Pisano considers the profitability of those that remain independent. But this is not necessarily the best measure, as these companies may be the ones that are doomed to fail. To be fair, he does warn that care should be taken when drawing conclusions from analyses of the output of the biotechnology sector.

The book asks a fundamental question about the financing of biotechnology: are the public equity markets the right place for biotechnology companies to raise their money? Pisano argues that structurally the answer is no, as investors can never have enough information on which to arrive at a valuation of an R&D programme in a revenue-less company.

In part, this depends on the reader's view of what an equity market exists to achieve. Mine is that equity markets are where companies seek capital. Not all companies succeed, with consequences for investors who, historically, accepted that potential large returns come with significant risk. Analysts would never recommend investing in a single biotechnology company, always presenting the industry as a portfolio play: nine companies will fail, but the tenth will deliver more than enough of a return to compensate. This highlights a transatlantic difference in the public equity markets that *Science Business*, with its focus on the United States, does not capture: the majority of investment in biotechnology on the London markets is institutional, a source of capital that can afford to invest in multiple companies and to take a long-term view.

More information is not the answer. Investors in biotechnology can already access accounts, scientific literature, conference proceedings and patents, and are updated by companies in some detail. This is more information than an investor in WalMart, say, might receive. Most investors are not, however, equipped to interpret these data.

Pisano concludes that the industry's failings can be corrected by better integration, better communication of information and better long-term learning. The evidence presented to support these conclusions is US-centric and therefore incomplete. Different strategies pursued elsewhere to encourage these solutions are not examined.

Science Business contains some useful facts but their value depends on the intended readership. For example, chapters 2 and 3 contain simple, if wordy, definitions of the many '-omics' in the biotechnology industry and a description of the drug development process. These are useful to those outside the industry but insiders can skip these chapters.

According to the sleeve notes, the book "provides clear prescriptions for companies, investors and policymakers seeking ways to improve the industry's performance". But does it? Read this book if you have an interest in the history of the commercialization of biotechnology, in how the industry works, and if you want to know what the issues facing the industry are. Just don't expect to find the answers it promises, merely a description of some of the options. ■

Keith Redpath is chief executive of Opal Drug Discovery, 41 Heriot Row, Edinburgh EH3 6ES, UK.

THEATRE

Here's looking at you

C. ROSEGG

The ninth annual season of the First Light Festival is under way in New York, offering new dramatic works exploring science and technology, commissioned by the Ensemble Studio Theatre and the Alfred P. Sloan Foundation. The centrepiece is the premiere of *Serendib*, a play by David Zellnik, inspired by two months the author spent with a team of primatologists studying toque macaques in Sri Lanka. The play focuses on scientists George Fischke and Anna Sunilagatte, who try to test the hypothesis that occasional violent takeovers of one troop of macaques by an outside alpha male adversely affects the troop's happiness. But money is short, so the scientists turn to two vaguely disreputable documentary film-makers for some much-needed publicity. The film-makers bring along Dmitri Ramsov, a geneticist who has no use for Fischke's woolly theories about the soulfulness of macaques. The stage is thus set for two mirror-image plays, with Fischke and Ramsov vying for supremacy and the affections of Sunilagatte, and the macaques under observation — puppets controlled by the actors — acting out a parallel drama. The film-makers' search for good television provides comic relief, but the play persists with its themes: the roles of observation and experiment in science, the distorting lens of media coverage, and the debate over the relative importance of heredity and environment in behaviour. It reaches a moving climax as Fischke and Ramsov deliver eloquent tributes to two of the macaques, their words illustrating the differences between humans and our primate cousins, and the powerful, if elusive, similarities. **Alan Packer**

http://ensemblestudiotheatre.org/index3.html

No firm promises

Science Business: The Promise, the Reality, and the Future of Biotech
by Gary P. Pisano
Harvard Business School Press: 2006.
256 pp. $29.95

Keith Redpath

I have an aversion to management books, and it is based on the way they are written. Why use a sentence when a chapter will do? Why use standard English when you can invent meaningless phrases? Gary Pisano suffers from this predilection in his book *Science Business*. He takes more than 70 pages to tell the reader that drug discovery is complex and risky, involving several disciplines and lots of money, and that biotechnology has simply increased the complexity. A further 30 pages are required to conclude that many commercialization strategies are being followed, but that it is too early to know which are right.

What Pisano does, however, which few have dared to do in print before, is the biotechnology equivalent of the boy who shouted out that the Emperor had no clothes. He poses the question: "can science be a business?" He concludes, on the basis of his research and the experience of biotechnology to date, that the answer is no, but that the industry still has time to exonerate itself, if only it can change.

Modern biotechnology was depicted at its inception as a huge danger to mankind. Genetic manipulation was either carried out behind greater security than was required for the reprocessing of nuclear waste, or it was banned altogether. We were told that mad scientists would release modified microbes that would wipe out mankind. The hysteria parallels that accompanying stem cells and nanotechnology some 30 years later and perhaps demonstrates one of Pisano's theses: that the industry fails to learn from its mistakes.

Biotechnology was soon repositioned as the saviour of mankind. No disease was immune to attack from the magic bullet of biotechnology. Shares soared. Investors cashed in and got rich. Many more companies formed. Universities saw ways of making money from research.

Despite that, Pisano contends that the industry has failed to deliver financially. He may be right, but his analysis is flawed. Many small start-up companies with promising drug candidates or technologies are acquired by larger companies before making a profit, delivering good returns for founders and investors. In deciding whether biotechnology has delivered financially, Pisano considers the profitability of those that remain independent. But this is not necessarily the best measure, as these companies may be the ones that are doomed to fail. To be fair, he does warn that care should be taken when drawing conclusions from analyses of the output of the biotechnology sector.

The book asks a fundamental question about the financing of biotechnology: are the public equity markets the right place for biotechnology companies to raise their money? Pisano argues that structurally the answer is no, as investors can never have enough information on which to arrive at a valuation of an R&D programme in a revenue-less company.

In part, this depends on the reader's view of what an equity market exists to achieve. Mine is that equity markets are where companies seek capital. Not all companies succeed, with consequences for investors who, historically, accepted that potential large returns come with significant risk. Analysts would never recommend investing in a single biotechnology company, always presenting the industry as a portfolio play: nine companies will fail, but the tenth will deliver more than enough of a return to compensate. This highlights a transatlantic difference in the public equity markets that *Science Business*, with its focus on the United States, does not capture: the majority of investment in biotechnology on the London markets is institutional, a source of capital that can afford to invest in multiple companies and to take a long-term view.

More information is not the answer. Investors in biotechnology can already access accounts, scientific literature, conference proceedings and patents, and are updated by companies in some detail. This is more information than an investor in WalMart, say, might receive. Most investors are not, however, equipped to interpret these data.

Pisano concludes that the industry's failings can be corrected by better integration, better communication of information and better long-term learning. The evidence presented to support these conclusions is US-centric and therefore incomplete. Different strategies pursued elsewhere to encourage these solutions are not examined.

Science Business contains some useful facts but their value depends on the intended readership. For example, chapters 2 and 3 contain simple, if wordy, definitions of the many '-omics' in the biotechnology industry and a description of the drug development process. These are useful to those outside the industry but insiders can skip these chapters.

According to the sleeve notes, the book "provides clear prescriptions for companies, investors and policymakers seeking ways to improve the industry's performance". But does it? Read this book if you have an interest in the history of the commercialization of biotechnology, in how the industry works, and if you want to know what the issues facing the industry are. Just don't expect to find the answers it promises, merely a description of some of the options. ■

Keith Redpath is chief executive of Opal Drug Discovery, 41 Heriot Row, Edinburgh EH3 6ES, UK.

Putting the pieces **together**

Rules of engagement

Complex engineered and biological systems share protocol-based architectures that make them robust and evolvable, but with hidden fragilities to rare perturbations.

John Doyle and Marie Csete

Chaos, fractals, random graphs and power laws inspire a popular view of complexity in which behaviours that are typically unpredictable and fragile 'emerge' from simple interconnections among like components. But applied to the study of highly evolved systems, this attractively simple view has led to widespread confusion. A different, more rewarding take on complexity focuses on organization, protocols and architecture, and includes the 'emergent' as an extreme special case within a much richer dynamical perspective.

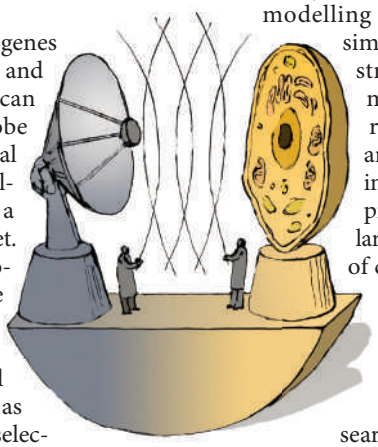
Engineers can learn from biology. Biological systems are robust and evolvable in the face of even large changes in environment and system components, yet can be extremely fragile to small perturbations. Such universally robust yet fragile (RYF) complexity is found wherever we look. Take the evolution of microbes into humans (robustness of lineages on long timescales) punctuated by mass extinctions (extreme fragility). Or diabetes and cancer, conditions resulting from faulty biological control mechanisms, normally so robust as to go unnoticed.

But RYF complexity is not confined to biology. The complexity of technology is exploding around us, but in ways that remain largely hidden. Modern institutions and technologies facilitate robustness and accelerate evolution, but also enable major catastrophes, from network crashes to climate change. Such RYF complexity presents a major challenge to engineers, physicians and, increasingly, scientists. Understanding RYF means understanding architecture — the most universal, high-level, persistent elements of organization — and protocols. Protocols define how diverse modules interact, and architecture defines how sets of protocols are organized.

So biologists can learn from engineering. The Internet is an obvious example of how a protocol-based architecture facilitates evolution and robustness. If you are reading this on the Internet, your laptop hardware (display, keyboard and so on) and software (web browser) both obey sets of protocols for exchanging signals and files. Subject to protocol-driven constraints, you can access an incredible diversity of hardware and software resources.

But it is the architecture of TCP/IP (Transmission Control and Internet Protocols) that is more fundamental. The hourglass protocol 'stack' has a thin, hidden 'waist' of universally shared feedback control (TCP/IP) between the visible upper (application software) and lower (hardware) layers. Roughly, IP controls the routes for packet flows and thus, available bandwidth. Applications split files into packets, and TCP controls their rates and guarantees delivery. This allows 'plug-and-play' between modules that obey shared protocols; any set of applications that 'talks' TCP can run transparently and robustly on any set of hardware that talks IP, accelerating the evolution of TCP/IP-based networks.

Similarly, microbial genes that talk transcription and translation protocols can move from one microbe to another by horizontal gene transfer, also accelerating evolution in a kind of bacterial internet. But as with the technological Internet, the newly acquired proteins work better when they can use additional shared protocols such as group transfers. Thus selection acting at the protocol level could evolve and preserve shared architecture, essentially evolving evolvability.



J. KAPUSTA/IMAGES.COM

All life and advanced technologies rely on protocol-based architectures. The evolvability of microbes and IP-based networks illustrates how dramatic, novel, dynamic changes on all scales of time and space can also be coherent, responsive, functional and adaptive. New genes and pathways, laptops and applications, even whole networks, can plug-and-play, as long as they obey protocols. Biologists can even swap gene sequences over the Internet in a kind of synthetic horizontal gene transfer.

Typical behaviour is fine-tuned with this elaborate control and thus appears boringly robust despite large internal and external perturbations. As a result, complexity and fragility are largely hidden, often revealed only by catastrophic failures. Because components come and go, control systems that reallocate network resources easily confer robustness to outright failures, whereas violations of protocols by even small random rewiring can be catastrophic. So programmed cell (or component) 'death' is a common strategy to prevent local failures from cascading system-wide.

The greatest fragility stemming from a reliance on protocols is that standardized interfaces and building blocks can be easily hijacked. So that which enables horizontal gene transfer, the web and email also aids viruses and other parasites. Large structured rearrangements can be tolerated, whereas small random or targeted changes that subtly violate protocols can be disastrous.

By contrast, in the popular view of complexity described at the beginning, modelling and analysis are both simplified because tuning, structure and details are minimized, as is environmental uncertainty; and superficial patterns in ensemble averages (not protocols) define modularity. An unfortunate clash of cultures arises because architecture-based RYF complexity is utterly bewildering when viewed from this popular perspective. But the search for a deep simplicity and unity remains a common goal.

Fortunately, our growing need for robust, evolvable technological networks means the tools for engineering architectures and protocols are becoming more accessible. These will bring rigour and relevance to the study of complexity generally, but not at the expense of structure and detail. Quite the contrary: both architectures and theories to study them are most successful when they facilitate rather than ignore the inclusion of domain-specific details and expertise. ■

John Doyle is at the California Institute of Technology, Pasadena, California 91125-8100, USA; Marie Csete is at Emory University, Atlanta, Georgia 30322, USA.

FURTHER READING
Doyle *et al. Proc. Natl Acad. Sci. USA* **102,** 14497–14502 (2005).
Moritz, M. A. *et al. Proc. Natl Acad. Sci. USA* **102,** 17912–17917 (2005).

For other essays in this series, see http://nature.com/nature/focus/arts/connections/index.html

CONNECTIONS

NEWS & VIEWS

PALAEOBOTANY

A tree without leaves

Brigitte Meyer-Berthaud and Anne-Laure Decombeix

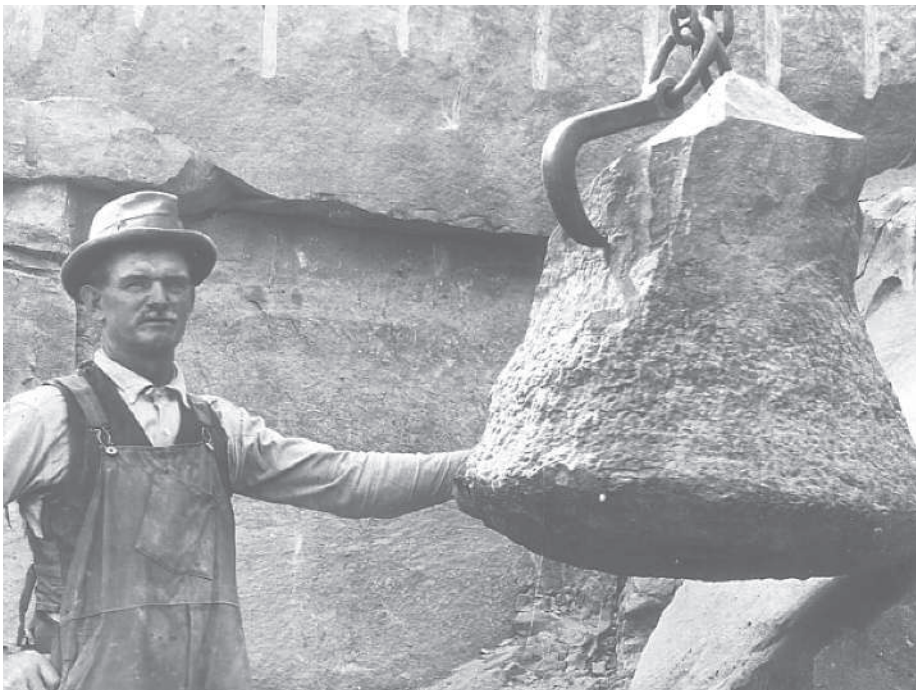
The puzzle presented by the famous stumps of Gilboa, New York, finds a solution in the discovery of two fossil specimens that allow the entire structure of these early trees to be reconstructed.

The Middle Devonian (397 million to 385 million years ago) was a notable time for the evolution of early land plants. The diversification of reproductive strategies, the advent of leaf precursors, and a tendency to increasing height all led to the rise of plants of modern appearance. One of the best locations for studying these processes is Gilboa, New York, which is internationally renowned for its wide range of organisms from the late Middle Devonian. These organisms include a rich variety of arthropods, for example, but also large, enigmatic stumps (Fig. 1) that in life formed extensive stands and were fossilized where they grew.

Since the discovery of these *Eospermatopteris* stumps in the nineteenth century, questions about their relationships to other plants and what kind of trees they supported have exercised generations of palaeobotanists. Evidence presented by Stein *et al.*[1] on page 904 of this issue not only answers these questions but also provides a new glimpse of the requirements necessary to make a tree.

Reconstructing entire fossil plants is an important step in assessing the patterns of plant diversification through time, and the roles that plants played in past environments. This task is challenging, because plants naturally shed parts of their body during their lifetime, and it is rarely achieved for structures such as trees, which have a complex architecture and an extended lifespan. However, the two exceptionally well-preserved specimens collected by Stein *et al.*[1] were sufficient to reconstruct the kind of tree that made the Gilboa forest (Fig. 2a, overleaf).

The two specimens partly overlap in size and morphology. One consists of a slender trunk more than 6 metres long, with an enlarged *Eospermatopteris*-type base, and covered with branch scars at the top. Slender 'coalified' strands extending downwards from the bottom of the stump are interpreted as roots. The other specimen is the top part of a trunk terminated by a crown of short and erect branches, and with branch scars at the bottom. Branches of the crown were divided digitately — that is, they produced several closely spaced branchlets flattened in a plane. The terminal appendages



JESSE EARL HYDE COLL., CASE WESTERN RES. UNIV., DEPT GEOL. SCI.

Figure 1 | An enigmatic stump. This photo (undated) was taken at the Riverside Quarry, Gilboa.

were three-dimensional and do not resemble leaves (organs with a flat blade of photosynthetic tissue). In the Devonian, this branching pattern characterizes the Pseudosporochnales, an extinct group traditionally considered to be closer to the ferns than to the seed plants and known to have included 3-metre-tall specimens about 392 million years ago[2]. This group was successful and occurred worldwide during the Middle Devonian[3].

One consequence of Stein and colleagues' discovery is that *Archaeopteris*, a close relative of seed plants that flourished in the Late Devonian (385 million to 359 million years ago), is no longer the oldest known modern tree (Fig. 2b). The conifer-like reconstruction of this 'progymnosperm' appears in most botany textbooks, and many of us are familiar with its woody stump, measuring up to 1.5 m in diameter, and which is suspected of having reached 40 m in height and to have been deeply anchored in the soil by an extensive root system[4,5]. *Archaeopteris* had large, perennial (long-lived) branches that were essential components of its architecture[6]. As it grew, *Archaeopteris* added more of these branches to its trunk, a process that contributed to its three-dimensional increase in size. Short-lived branches produced at the periphery were leafy.

The Gilboa tree differs from *Archaeopteris* in having a trunk of more moderate size, in lacking perennial branches, and in possessing a limited root system that apparently consisted of many roots of similar size. The short, disposable branches of the crown were shed when senescent, and were continuously replaced by new branches at the top of the trunk. Unlike *Archaeopteris*, the root and branch systems in the Gilboa tree did not increase much in extent during growth. In this, the tree resembles the tree-ferns, and the cycads and palms with single trunks that occur today. But unlike these, it was unable to make planate leaves; instead, it seems that photosynthesis was carried out by the periodically shed branches of the crown and, more

50 YEARS AGO

The report of Dr. C. P. Haskins, president of the Carnegie Institution of Washington ...comments on the slight change during half a century in the structure and basic orientation of the Institution; although when the Institution was founded, the idea, let alone the practice, of scientific investigation in the United States was almost unknown, except as an integral but subsidiary part of university work, and an editorial in *Science* during 1903 expressed grave doubts as to the function or future of an organization devoted to such an idea...[Dr Haskins] believes, despite the radical changes in the American scientific scene, that this conception of the Institution has stood the test of half a century and is as relevant today. The Institution has been able to play an important part in technological change in times of national emergency; and today...continues to pursue its aims of research towards the same goals and in essentially the same manner as it has throughout its working life.

From *Nature* 20 April 1957

100 YEARS AGO

Apparently, the British government is indifferent to any increase of facilities for the advancement of knowledge, for it makes no attempt to show active interest in organisations and institutions concerned with science and higher education. The Carnegie Institute at Pittsburg was dedicated last week in the presence of a large and distinguished company, but neither the British ambassador nor any member of the British embassy was present...The German Emperor was represented by a special commission of six members of the highest rank; France and Italy were also represented...The omission is only another instance of the failure of British statesmen to understand the significance of anything relating to science or progressive learning.

From *Nature* 18 April 1907

50 & 100 YEARS AGO

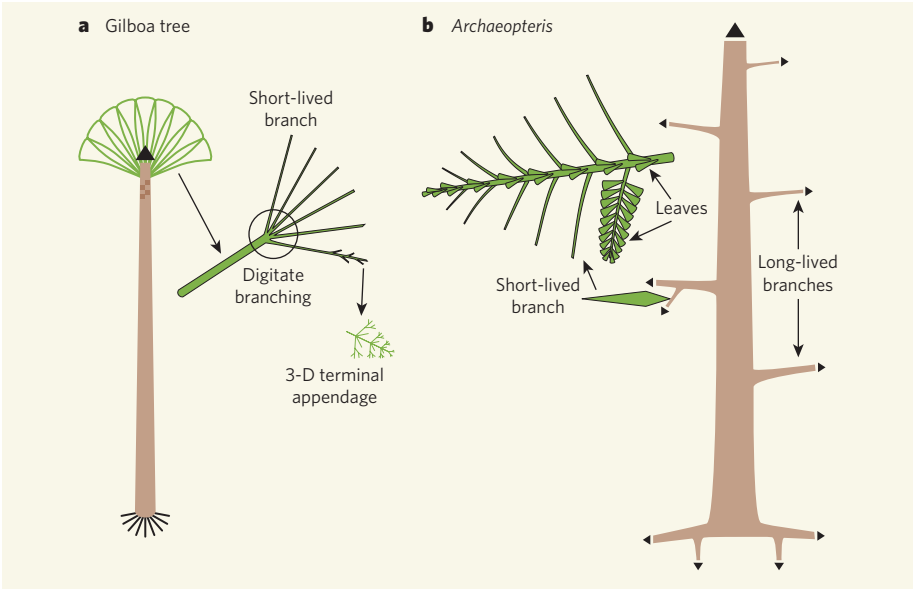


Figure 2 | Two types of Devonian tree. **a**, The newly described Gilboa tree[1], a member of the Pseudosporochnales (Fig. 3), had no leaves and a limited root system, and displayed an economical strategy whereby a single, long-lived organ — the trunk — grew vertically. **b**, In contrast, *Archaeopteris* possessed leafy twigs, and had long-lived roots and branches that grew at the same time as the trunk. Photosynthetic organs are shown in green; black triangles indicate long-lived organs.

specifically, by their three-dimensional terminal appendages.

One benefit often assumed for taller plants is their enhanced ability to capture light. Ten years ago, Niklas[7] simulated the architecture of early land plants and tested their efficiency in performing several essential functions. The Gilboa tree fits closely with the morphology that optimizes two functions, mechanical stability and reproduction. But the reduced surface area of its crown was not optimal for light interception.

Two contrasting ways of making trees evolved during the Devonian (Figs 2a, b; Fig. 3). The way represented by *Archaeopteris*, and by most extant trees of temperate and tropical areas, requires a complex machinery of tissues and organs to achieve growth in all spatial directions and to build the larger body sizes recorded in the plant kingdom. The Gilboa tree represents an economical alternative where, beyond the necessary investment in spores to ensure reproduction, the products of photosynthesis were mainly devoted to vertical growth of the trunk. The new specimens from New York[1] show that the first giants in the history of the land plants achieved the tree habit and significant biomass despite their inability to construct optimal photosynthetic structures, such as leaves or horizontal branches, and despite not building an extensive root system. ■

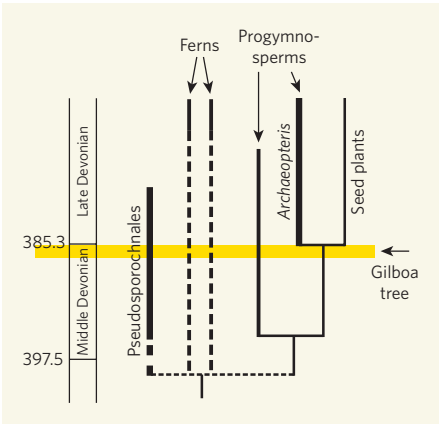


Figure 3 | Trees in time. Two contrasting ways of making trees, evident in the fossils of the Gilboa tree and of *Archaeopteris*, evolved in the Devonian but are still found today. The Gilboa tree is a member of the extinct group, Pseudosporochnales. *Archaeopteris* is a member of the progymnosperms, extinct relatives of the seed plants[6,8]. Timescale is millions of years ago. Many representatives of ferns and seed plants exist today, the latter being by far the main constituent of the current terrestrial flora.

Brigitte Meyer-Berthaud (CNRS) and Anne-Laure Decombeix (Université Montpellier 2) are in the Unité Mixte de Recherche Botanique et Bioinformatique de l'Architecture des Plantes (AMAP), CIRAD, 34398 Montpellier Cedex 5, France.
e-mails: meyerberthaud@cirad.fr; anne-laure.decombeix@cirad.fr

1. Stein, W. E., Mannolini, F., Hernick, L. V., Landing, E. & Berry, C. M. *Nature* **446,** 904–907 (2007).
2. Berry, C. M. & Fairon-Demaret, M. *Int. J. Plant Sci.* **163,** 699–713 (2002).
3. Berry, C. M. & Fairon-Demaret, M. in *Plants Invade the Land: Evolutionary and Environmental Perspectives* (eds Gensel, P. G. & Edwards, D.) 120–139 (Columbia Univ. Press, New York, 2001).
4. Beck, C. B. & Wight, D. C. in *Origin and Evolution of Gymnosperms* (ed. Beck, C. B.) 1–84 (Columbia Univ. Press, New York, 1988).
5. Scheckler, S. E. in *Palaeobiology II* (eds Briggs, D. E. G. & Crowther, P.) 67–71 (Blackwell Science, Oxford, 2001).
6. Meyer-Berthaud, B. *et al. Nature* **398,** 700–701 (1999).
7. Niklas, K. J. *Am. J. Bot.* **84,** 16–25 (1997).
8. Rothwell, G. W. & Nixon, K. C. *Int. J. Plant Sci.* **167,** 737–749 (2006).

50 YEARS AGO

The report of Dr. C. P. Haskins, president of the Carnegie Institution of Washington ...comments on the slight change during half a century in the structure and basic orientation of the Institution; although when the Institution was founded, the idea, let alone the practice, of scientific investigation in the United States was almost unknown, except as an integral but subsidiary part of university work, and an editorial in *Science* during 1903 expressed grave doubts as to the function or future of an organization devoted to such an idea...[Dr Haskins] believes, despite the radical changes in the American scientific scene, that this conception of the Institution has stood the test of half a century and is as relevant today. The Institution has been able to play an important part in technological change in times of national emergency; and today...continues to pursue its aims of research towards the same goals and in essentially the same manner as it has throughout its working life.

From *Nature* 20 April 1957

100 YEARS AGO

Apparently, the British government is indifferent to any increase of facilities for the advancement of knowledge, for it makes no attempt to show active interest in organisations and institutions concerned with science and higher education. The Carnegie Institute at Pittsburg was dedicated last week in the presence of a large and distinguished company, but neither the British ambassador nor any member of the British embassy was present...The German Emperor was represented by a special commission of six members of the highest rank; France and Italy were also represented...The omission is only another instance of the failure of British statesmen to understand the significance of anything relating to science or progressive learning.

From *Nature* 18 April 1907

50 & 100 YEARS AGO

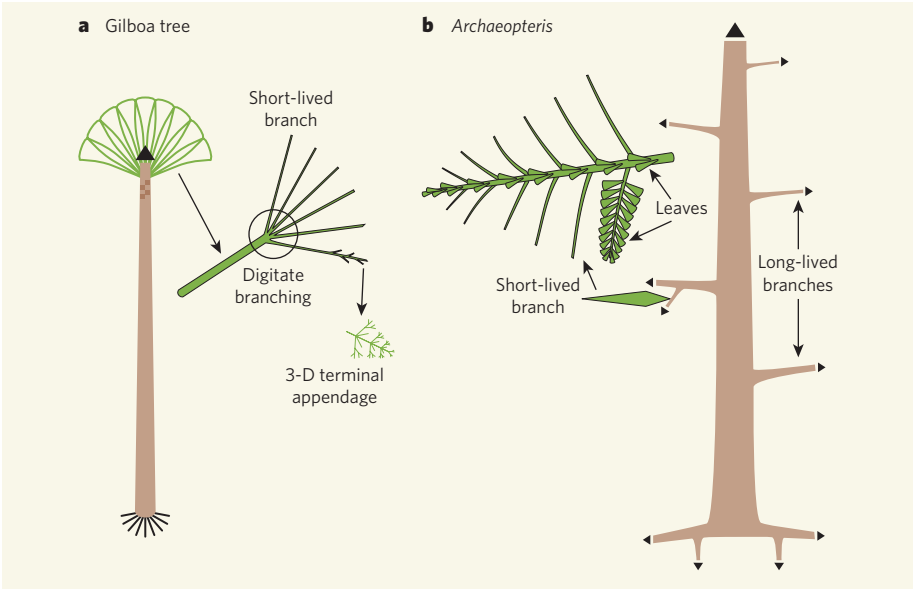


Figure 2 | Two types of Devonian tree. **a**, The newly described Gilboa tree[1], a member of the Pseudosporochnales (Fig. 3), had no leaves and a limited root system, and displayed an economical strategy whereby a single, long-lived organ — the trunk — grew vertically. **b**, In contrast, *Archaeopteris* possessed leafy twigs, and had long-lived roots and branches that grew at the same time as the trunk. Photosynthetic organs are shown in green; black triangles indicate long-lived organs.

specifically, by their three-dimensional terminal appendages.

One benefit often assumed for taller plants is their enhanced ability to capture light. Ten years ago, Niklas[7] simulated the architecture of early land plants and tested their efficiency in performing several essential functions. The Gilboa tree fits closely with the morphology that optimizes two functions, mechanical stability and reproduction. But the reduced surface area of its crown was not optimal for light interception.

Two contrasting ways of making trees evolved during the Devonian (Figs 2a, b; Fig. 3). The way represented by *Archaeopteris*, and by most extant trees of temperate and tropical areas, requires a complex machinery of tissues and organs to achieve growth in all spatial directions and to build the larger body sizes recorded in the plant kingdom. The Gilboa tree represents an economical alternative where, beyond the necessary investment in spores to ensure reproduction, the products of photosynthesis were mainly devoted to vertical growth of the trunk. The new specimens from New York[1] show that the first giants in the history of the land plants achieved the tree habit and significant biomass despite their inability to construct optimal photosynthetic structures, such as leaves or horizontal branches, and despite not building an extensive root system. ■

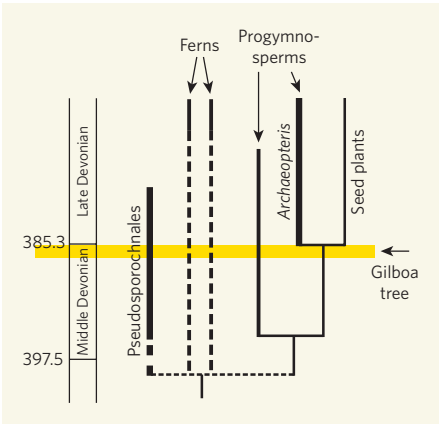


Figure 3 | Trees in time. Two contrasting ways of making trees, evident in the fossils of the Gilboa tree and of *Archaeopteris*, evolved in the Devonian but are still found today. The Gilboa tree is a member of the extinct group, Pseudosporochnales. *Archaeopteris* is a member of the progymnosperms, extinct relatives of the seed plants[6,8]. Timescale is millions of years ago. Many representatives of ferns and seed plants exist today, the latter being by far the main constituent of the current terrestrial flora.

Brigitte Meyer-Berthaud (CNRS) and Anne-Laure Decombeix (Université Montpellier 2) are in the Unité Mixte de Recherche Botanique et Bioinformatique de l'Architecture des Plantes (AMAP), CIRAD, 34398 Montpellier Cedex 5, France.
e-mails: meyerberthaud@cirad.fr; anne-laure.decombeix@cirad.fr

1. Stein, W. E., Mannolini, F., Hernick, L. V., Landing, E. & Berry, C. M. *Nature* **446,** 904–907 (2007).
2. Berry, C. M. & Fairon-Demaret, M. *Int. J. Plant Sci.* **163,** 699–713 (2002).
3. Berry, C. M. & Fairon-Demaret, M. in *Plants Invade the Land: Evolutionary and Environmental Perspectives* (eds Gensel, P. G. & Edwards, D.) 120–139 (Columbia Univ. Press, New York, 2001).
4. Beck, C. B. & Wight, D. C. in *Origin and Evolution of Gymnosperms* (ed. Beck, C. B.) 1–84 (Columbia Univ. Press, New York, 1988).
5. Scheckler, S. E. in *Palaeobiology II* (eds Briggs, D. E. G. & Crowther, P.) 67–71 (Blackwell Science, Oxford, 2001).
6. Meyer-Berthaud, B. *et al. Nature* **398,** 700–701 (1999).
7. Niklas, K. J. *Am. J. Bot.* **84,** 16–25 (1997).
8. Rothwell, G. W. & Nixon, K. C. *Int. J. Plant Sci.* **167,** 737–749 (2006).

EARTH SCIENCE

When crust is bred

Don Porcelli

An analysis of the distribution of helium-isotope ratios in oceanic extrusions from Earth's mantle seems to establish a connection with the spread of ages in continental crust. What mechanism might underlie this?

After decades of intensive study, fundamental questions remain about the chemical composition and evolution of Earth's interior. On page 900 of this issue, Parman[1] reports on an analysis of helium-isotope ratios in basalts found in oceanic islands and derived from Earth's mantle. He finds an intriguing correlation between the distribution of helium ratios — a chemical indicator of mantle history — and the age-distribution of rocks in Earth's continental crust.

Helium escapes from Earth's mantle at zones where melting occurs. Gases such as helium enter into melts that rise to the surface and are lost to the atmosphere during volcanic eruption — a process known as mantle degassing. The lighter isotope, ^{3}He, is almost exclusively primordial (it has been present unchanged since Earth first formed), and so is largely lost from the mantle through the degassing process. The heavier ^{4}He isotope, however, can be replenished: it is continuously produced by decay of the radioactive elements uranium and thorium, which are also present in the mantle. The ^{4}He/^{3}He ratio, therefore, reflects the long-term average U/^{3}He ratio of the rock source.

Rock from the volcanoes that form Earth's 40,000-km-long mid-ocean ridge system, which lines constructive tectonic-plate boundaries, have a relatively uniform, high ^{4}He/^{3}He ratio of about 9×10^4. This seems to indicate that these rocks, known as mid-ocean-ridge basalts (MORBs), erupted from a homogeneous upper mantle that undergoes continual, though incomplete, degassing. Remarkably, however, many ocean-island basalts (OIBs) have a range of ^{4}He/^{3}He values up to five times lower than the MORB values. OIBs stem from volcanoes in ocean islands away from tectonic-plate margins, and their ratios seem to reflect mantle-source regions with lower U/^{3}He ratios and different degassing histories. This had been widely interpreted as reflecting the involvement of an isolated, separately convecting, gas-rich region deep in the mantle. This deep mantle was spared from degassing at volcanic zones — the fate of the MORB mantle — and plumes rising from it delivered material to ocean islands.

However, recent geophysical evidence suggests that subducting ocean crust passes into the deep mantle, implying that such an isolated, deep layer does not exist. As an alternative, it was proposed that low source U/^{3}He ratios can be generated continually by melting within the upper mantle[2]. Both helium and uranium are largely removed by mantle melting and degassing. But if uranium were more effectively removed, the residue would maintain a lower ^{4}He/^{3}He ratio than the rest of the mantle, which would continue to evolve to higher ratios through the addition of ^{4}He from the decay of uranium. Laboratory experiments[3,4] show that uranium may indeed be more efficiently removed, and that the lowering of the U/^{3}He ratio is most pronounced if large amounts of melting occur. But more direct evidence is required to prove that the low ^{4}He/^{3}He ratios seen in OIBs are actually generated in this way.

Parman[1] argues that a direct link can be made between helium compositions in OIBs and the times of mantle depletion through continental-crust extraction. Histories of continental-crust growth have been constructed from the age distributions of zircons, minerals that form as new crust is created and that remain resistant to alteration[5,6]. When OIB ^{4}He/^{3}He ratios are plotted — on a scale from the high MORB values to the low primordial values — and are compared with the distribution of zircon ages spanning the lifetime of Earth, the distribution of peaks is the same (Fig. 2 on page 902).

Such a correlation requires a connection between the formation of new crust and mantle composition. Assuming that the ^{4}He/^{3}He ratio of the mantle has evolved linearly, the helium-isotope compositions of OIBs are directly proportional to the time at which their sources were formed, and so can, in turn, be directly compared to crustal growth ages. Parman suggests that, when a section of crust is formed by partial melting of the mantle, a correspondingly depleted section of mantle is created that preserves the distinctive ^{4}He/^{3}He ratio of the bulk mantle at that time. This record then appears, through the activity of mantle plumes, in OIBs (Fig. 1). The correlation supports both the idea that there have been different periods of peak crustal growth rate (and so mantle melting), and the idea that the source of OIBs includes material that was strongly depleted by melting in the past. It also provides an explanation for the range of helium-isotope signatures found in the mantle.

But several important questions surround this theory. First, do the peaks in the distribution of the helium-isotope data really reflect the distribution of compositions in the mantle? Studies generally make no attempt to determine the volume of basalt that each isotope composition represents, and the distribution of the data is expected to reflect factors such as accessibility to sampling areas, variations in surface exposures and other sampling biases.

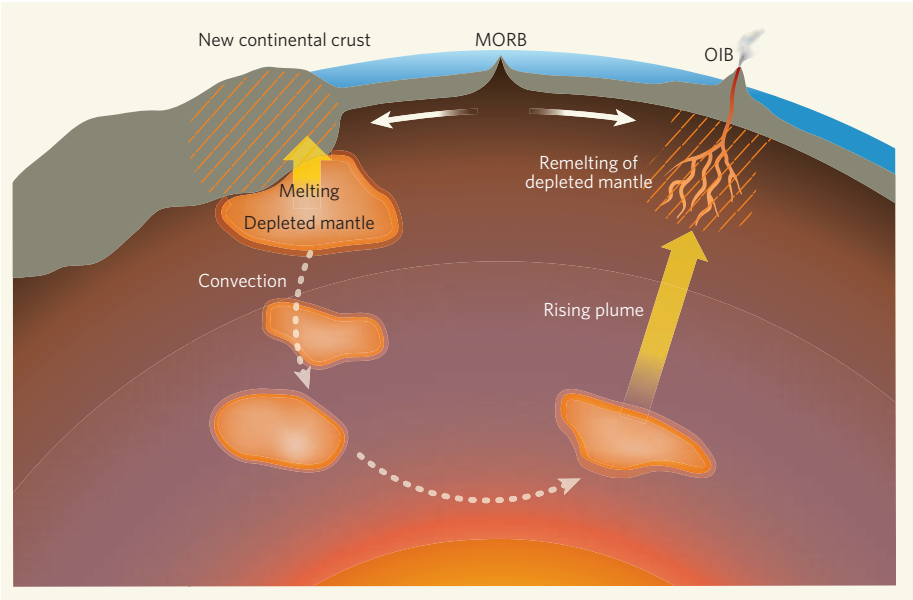


Figure 1 | Island sources for crustal growth. Parman[1] describes a possible connection between crustal growth and the development of sources for ocean-island basalts (OIBs). The continuous loss of helium in its two isotopic forms, ^{3}He and ^{4}He, to the atmosphere through degassing, together with the replenishment of ^{4}He by radioactive decay of uranium and thorium, have brought about a progressive increase in the ^{4}He/^{3}He ratio of Earth's mantle through history. But whenever a portion of the mantle melts to form continental crust, it is left depleted in uranium and thorium relative to helium. Its ^{4}He/^{3}He ratio cannot therefore increase further, and these mantle regions preserve the helium-isotope ratio of the mantle at that time. The material from these regions circulates back into the bulk mantle and later rises up in mantle plumes to become the source regions for OIBs. This would explain the discrepancy between these rocks' helium ratio and that of basalts from mid-ocean ridges (MORBs), which mirror the bulk, mantle composition.

In addition, mixing between mantle components can generate a range of basalt isotope compositions more varied than that seen in the mantle.

Second, can depleted material really impart distinctive helium-isotope signals into OIBs? Both helium and uranium are strongly concentrated into mantle melts. This means that, although low U/^{3}He ratios may be left behind during melting, there is very little helium left at all, especially if large amounts of melting occur. The helium-isotope signature of the depleted material is likely to be hidden by that of various other mantle components, such as subducted material and surrounding, less depleted bulk mantle that clearly also contribute to the source of OIBs. Furthermore, other signals that would characterize the depleted regions of mantle suggested by Parman[1] as the origin of OIBs, including neodymium, lead and hafnium isotopes, are typically not found in OIBs.

Third, how would such a mantle work? Mantle domains now seen in ocean islands would have been generated throughout Earth's history, and so must have been stored where mixing would not destroy them. This is a particular problem for the largest hotspots of Hawaii and Iceland, which have the lowest ^{4}He/^{3}He ratios and so would need a source as old as 4 billion years. Although it is tempting to picture separate blobs circulating in the mantle like raisins in a pudding, these would probably be destroyed by mixing[7]. And questions of how this material can be preferentially incorporated into plumes beneath ocean islands would remain. Even more perplexing, how can individual locations such as Hawaii and Iceland have basalts derived from a range of these sources, as is implied by the spread of helium data within these regions?

Overall, the picture suggested by Parman provides an elegant explanation for a fundamental characteristic of ocean-island rocks. Although there are still considerable reasons to be sceptical, the many points of correspondence between peaks in zircon ages in the crust and helium isotopes in the mantle seem to be more than a coincidence. If this observation survives further scrutiny and data collection, it will certainly be important for understanding some of the most distinctive volcanic features on Earth. ■

Don Porcelli is in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.
e-mail: don.porcelli@earth.ox.ac.uk

1. Parman, S. W. *Nature* **446,** 900–903 (2007).
2. Graham, D. W., Lupton, F., Albarède, F. & Condomines, M. *Nature* **347,** 545–548 (1990).
3. Parman, S. W., Kurz, M. D., Hart, S. R. & Grove, T. L. *Nature* **437,** 1140–1143 (2005).
4. Heber, V. S., Brooker, R. A., Kelley, S. P. & Wood, B. J. *Geochim. Cosmochim. Acta* **71,** 1041–1061 (2007).
5. Condie, K. C. *Earth Planet. Sci. Lett.* **163,** 97–108 (1998).
6. Kemp, A. I. S., Hawkesworth, C. J., Paterson, B. A. & Kinny, P. D. *Nature* **439,** 580–583 (2006).
7. van Keken, P. E., Hauri, E. H. & Ballentine, C. J. *Annu. Rev. Earth Planet. Sci.* **30,** 493–525 (2002).

RNA SILENCING

Genomic defence with a slice of pi

Phillip D. Zamore

In fruit flies, a few very large genes generate the small RNAs that silence parasitic DNA elements. These RNAs might also participate in an amplification circuit that increases their potency.

Nearly half of the human genome and a third of the fruit fly's consists of selfish elements called transposons, 'jumping genes' that insert themselves into new locations, mutating other genes and damaging chromosomes. These molecular parasites include simple nucleotide repeats and virus-like elements that have colonized genomes throughout evolution. In response, animals have evolved complex mechanisms to silence transposons. The fruit fly *Drosophila melanogaster* silences its transposons by a mechanism that uses small RNA 'guides'. Reporting in *Cell*[1] and *Science*[2], two research groups propose an unanticipated explanation for the production and amplification of these small, silencing RNAs.

The best known silencing pathway guided by small RNAs is RNA interference (RNAi), in which small interfering RNAs (siRNAs) trigger an immune response that defends plants and animals against viruses by destroying complementary RNA sequences[3,4]. The siRNAs bind to members of the Argonaute family of proteins — molecular 'scissors' that use the small RNA guide to bind to and cut a second RNA molecule, the 'target'. The use of RNA guides rather than protein antibodies is a key difference between the sequence-based RNAi defence pathway and our adaptive immune responses, which recognize viral proteins. The enzyme Dicer produces siRNAs by cutting long, double-stranded RNA into smaller pieces about 21 nucleotides long. One strand of each siRNA is then loaded onto an Argonaute protein, generating a RNA–protein complex that binds to a viral RNA target by base-pairing and cuts it.

Like viruses, transposons are parasites, but, unlike viruses, they are primarily transmitted by inheritance, rather than through infection. To produce enzymes that facilitate their jumping to a new location in the genome, transposons must first be copied into mRNA. The evolution of small-RNA-guided silencing mechanisms, which prevent copying of transposons into RNA sequences, allows higher organisms to defend their genes against transposons.

In some higher organisms, the RNAi pathway provides a crucial defence mechanism against transposons[5–7]. But for fruit flies the genome is guarded by the Piwi-associated interfering RNA (piRNA) pathway[8]. At the heart of this pathway lies a specialized set of Argonaute proteins produced in the germ cells. Flies have five Argonaute proteins: Ago1, which uses small RNAs called microRNAs to regulate gene expression; Ago2, which uses siRNAs to fight viral infection; and three closely related Piwi proteins — Piwi, Aubergine and Argonaute3 (Ago3) — which use piRNAs to silence transposons and other parasitic DNAs. In flies, piRNAs are also called repeat-associated siRNAs, or rasiRNAs.

The Hannon[1] and Siomi[2] laboratories separately set out to identify small RNAs bound to each of the three fly Piwi proteins, in the hope that their sequences would reveal how rasiRNAs are made and function. Their findings are simply spectacular, suggesting that rasiRNAs arise from a small number of trigger loci — huge 'genes' that produce small RNAs against many selfish genetic elements — and that they are amplified through reciprocal cycles of cleavage by pairs of Piwi proteins (Fig. 1, overleaf).

To identify the RNAs bound to the three *Drosophila* Piwi proteins, each laboratory used antibodies to purify Piwi, Aubergine and Ago3 from the fly's ovaries, along with their associated rasiRNAs. Hannon and colleagues[1] identified 60,691 different rasiRNAs. It is difficult to pinpoint the site of origin of any one rasiRNA to a single genomic location, because nearly identical copies of specific transposons litter the entire genome. However, the set of rasiRNA sequences identified by these authors was so large that around 12,000 could be assigned to unique sites. These sites formed just 142 clusters, with a single cluster on the right arm of chromosome 2 comprising about 21% of all the unique rasiRNAs.

Another cluster corresponded to a genetic locus called *flamenco* on the X chromosome, which was previously shown to repress the jumping of the *gypsy*, *Idefix* and *ZAM* transposons. Since its discovery[9], *flamenco* has posed a puzzle because no protein-encoding gene resides at this locus. The new results indicate that *flamenco*, instead of producing a protein, is the source of rasiRNAs that target multiple types of transposon. These rasiRNAs may come from an enormous precursor RNA molecule, as mutations lying at the beginning of the locus disrupt *flamenco*-derived rasiRNAs some 168,000 nucleotides away. Loss of *flamenco* function activates transposons, such as *gypsy*, that lie within it, although most *gypsy* transposons reside outside this locus. However, it has no effect on transposons that are not found in *flamenco*. Thus, *flamenco* seems to have evolved as a master regulator of *gypsy*, *Idefix*

In addition, mixing between mantle components can generate a range of basalt isotope compositions more varied than that seen in the mantle.

Second, can depleted material really impart distinctive helium-isotope signals into OIBs? Both helium and uranium are strongly concentrated into mantle melts. This means that, although low U/^{3}He ratios may be left behind during melting, there is very little helium left at all, especially if large amounts of melting occur. The helium-isotope signature of the depleted material is likely to be hidden by that of various other mantle components, such as subducted material and surrounding, less depleted bulk mantle that clearly also contribute to the source of OIBs. Furthermore, other signals that would characterize the depleted regions of mantle suggested by Parman[1] as the origin of OIBs, including neodymium, lead and hafnium isotopes, are typically not found in OIBs.

Third, how would such a mantle work? Mantle domains now seen in ocean islands would have been generated throughout Earth's history, and so must have been stored where mixing would not destroy them. This is a particular problem for the largest hotspots of Hawaii and Iceland, which have the lowest ^{4}He/^{3}He ratios and so would need a source as old as 4 billion years. Although it is tempting to picture separate blobs circulating in the mantle like raisins in a pudding, these would probably be destroyed by mixing[7]. And questions of how this material can be preferentially incorporated into plumes beneath ocean islands would remain. Even more perplexing, how can individual locations such as Hawaii and Iceland have basalts derived from a range of these sources, as is implied by the spread of helium data within these regions?

Overall, the picture suggested by Parman provides an elegant explanation for a fundamental characteristic of ocean-island rocks. Although there are still considerable reasons to be sceptical, the many points of correspondence between peaks in zircon ages in the crust and helium isotopes in the mantle seem to be more than a coincidence. If this observation survives further scrutiny and data collection, it will certainly be important for understanding some of the most distinctive volcanic features on Earth. ■

Don Porcelli is in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.
e-mail: don.porcelli@earth.ox.ac.uk

1. Parman, S. W. *Nature* **446,** 900–903 (2007).
2. Graham, D. W., Lupton, F., Albarède, F. & Condomines, M. *Nature* **347,** 545–548 (1990).
3. Parman, S. W., Kurz, M. D., Hart, S. R. & Grove, T. L. *Nature* **437,** 1140–1143 (2005).
4. Heber, V. S., Brooker, R. A., Kelley, S. P. & Wood, B. J. *Geochim. Cosmochim. Acta* **71,** 1041–1061 (2007).
5. Condie, K. C. *Earth Planet. Sci. Lett.* **163,** 97–108 (1998).
6. Kemp, A. I. S., Hawkesworth, C. J., Paterson, B. A. & Kinny, P. D. *Nature* **439,** 580–583 (2006).
7. van Keken, P. E., Hauri, E. H. & Ballentine, C. J. *Annu. Rev. Earth Planet. Sci.* **30,** 493–525 (2002).

RNA SILENCING

Genomic defence with a slice of pi

Phillip D. Zamore

In fruit flies, a few very large genes generate the small RNAs that silence parasitic DNA elements. These RNAs might also participate in an amplification circuit that increases their potency.

Nearly half of the human genome and a third of the fruit fly's consists of selfish elements called transposons, 'jumping genes' that insert themselves into new locations, mutating other genes and damaging chromosomes. These molecular parasites include simple nucleotide repeats and virus-like elements that have colonized genomes throughout evolution. In response, animals have evolved complex mechanisms to silence transposons. The fruit fly *Drosophila melanogaster* silences its transposons by a mechanism that uses small RNA 'guides'. Reporting in *Cell*[1] and *Science*[2], two research groups propose an unanticipated explanation for the production and amplification of these small, silencing RNAs.

The best known silencing pathway guided by small RNAs is RNA interference (RNAi), in which small interfering RNAs (siRNAs) trigger an immune response that defends plants and animals against viruses by destroying complementary RNA sequences[3,4]. The siRNAs bind to members of the Argonaute family of proteins — molecular 'scissors' that use the small RNA guide to bind to and cut a second RNA molecule, the 'target'. The use of RNA guides rather than protein antibodies is a key difference between the sequence-based RNAi defence pathway and our adaptive immune responses, which recognize viral proteins. The enzyme Dicer produces siRNAs by cutting long, double-stranded RNA into smaller pieces about 21 nucleotides long. One strand of each siRNA is then loaded onto an Argonaute protein, generating a RNA–protein complex that binds to a viral RNA target by base-pairing and cuts it.

Like viruses, transposons are parasites, but, unlike viruses, they are primarily transmitted by inheritance, rather than through infection. To produce enzymes that facilitate their jumping to a new location in the genome, transposons must first be copied into mRNA. The evolution of small-RNA-guided silencing mechanisms, which prevent copying of transposons into RNA sequences, allows higher organisms to defend their genes against transposons.

In some higher organisms, the RNAi pathway provides a crucial defence mechanism against transposons[5–7]. But for fruit flies the genome is guarded by the Piwi-associated interfering RNA (piRNA) pathway[8]. At the heart of this pathway lies a specialized set of Argonaute proteins produced in the germ cells. Flies have five Argonaute proteins: Ago1, which uses small RNAs called microRNAs to regulate gene expression; Ago2, which uses siRNAs to fight viral infection; and three closely related Piwi proteins — Piwi, Aubergine and Argonaute3 (Ago3) — which use piRNAs to silence transposons and other parasitic DNAs. In flies, piRNAs are also called repeat-associated siRNAs, or rasiRNAs.

The Hannon[1] and Siomi[2] laboratories separately set out to identify small RNAs bound to each of the three fly Piwi proteins, in the hope that their sequences would reveal how rasiRNAs are made and function. Their findings are simply spectacular, suggesting that rasiRNAs arise from a small number of trigger loci — huge 'genes' that produce small RNAs against many selfish genetic elements — and that they are amplified through reciprocal cycles of cleavage by pairs of Piwi proteins (Fig. 1, overleaf).

To identify the RNAs bound to the three *Drosophila* Piwi proteins, each laboratory used antibodies to purify Piwi, Aubergine and Ago3 from the fly's ovaries, along with their associated rasiRNAs. Hannon and colleagues[1] identified 60,691 different rasiRNAs. It is difficult to pinpoint the site of origin of any one rasiRNA to a single genomic location, because nearly identical copies of specific transposons litter the entire genome. However, the set of rasiRNA sequences identified by these authors was so large that around 12,000 could be assigned to unique sites. These sites formed just 142 clusters, with a single cluster on the right arm of chromosome 2 comprising about 21% of all the unique rasiRNAs.

Another cluster corresponded to a genetic locus called *flamenco* on the X chromosome, which was previously shown to repress the jumping of the *gypsy*, *Idefix* and *ZAM* transposons. Since its discovery[9], *flamenco* has posed a puzzle because no protein-encoding gene resides at this locus. The new results indicate that *flamenco*, instead of producing a protein, is the source of rasiRNAs that target multiple types of transposon. These rasiRNAs may come from an enormous precursor RNA molecule, as mutations lying at the beginning of the locus disrupt *flamenco*-derived rasiRNAs some 168,000 nucleotides away. Loss of *flamenco* function activates transposons, such as *gypsy*, that lie within it, although most *gypsy* transposons reside outside this locus. However, it has no effect on transposons that are not found in *flamenco*. Thus, *flamenco* seems to have evolved as a master regulator of *gypsy*, *Idefix*

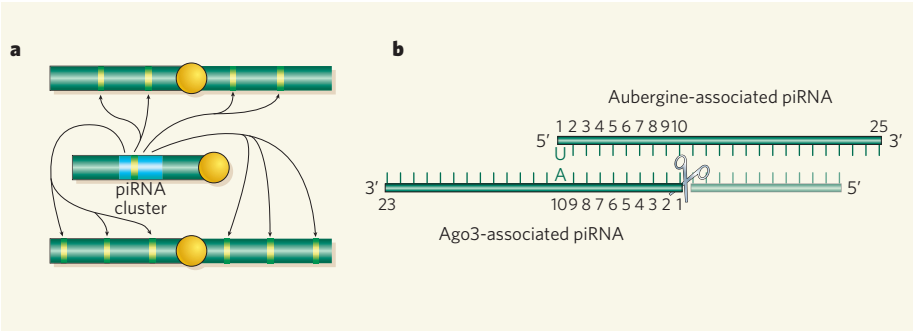


Figure 1 | piRNA-mediated silencing of transposons. **a**, A few trigger loci generate piRNAs, which target transposons (yellow) elsewhere in the genome and prevent their transcription. **b**, Complementary binding of Aubergine- and Ago3-associated piRNA sequences results in their amplification, ensuring efficient silencing of transposons.

and *ZAM* transposons and is the primary, if not the sole, source of rasiRNAs against these selfish elements.

Most rasiRNAs correspond to the antisense strand of transposons and can therefore bind to, and presumably destroy, the RNA transcripts of transposons[1,8]. These antisense rasiRNAs bind to Piwi and Aubergine. The two new papers[1,2] report that the small RNAs bound to Ago3 are nearly all of the sense orientation. Of the 353 Ago3-associated rasiRNAs identified by Siomi's group[2], the first 10 nucleotides of 16 sequences could be paired with rasiRNAs bound to Aubergine. And more than 11,200 (48%) of the Ago3-associated sense rasiRNAs identified by Hannon's group[1] could form an offset couple with at least one antisense rasiRNA.

A 10-nucleotide offset between the beginning of a small RNA guide and a second RNA molecule has a special meaning. Argonaute proteins cut target RNAs by measuring 10 nucleotides from the beginning of their RNA guide to the site of cleavage on their target. Such a pairing scheme suggests that the starting nucleotide of each antisense rasiRNA is defined by a cut that is guided by a corresponding sense rasiRNA. Reinforcing this view, nearly all the antisense rasiRNAs begin with the nucleotide U, whereas the sense rasiRNAs show no bias for beginning with U, A, C or G. Instead, the tenth nucleotide of the sense rasiRNAs was almost always A, which would allow it to pair with the first nucleotide — U — of an Aubergine- or Piwi-bound antisense rasiRNA (Fig. 1b).

Imagine, then, that a mother fly protects her offspring by providing her developing eggs with some of her Aubergine- and Piwi-bound antisense rasiRNAs. These could then generate sense rasiRNAs by cleaving the RNA transcripts of transposons, thereby simultaneously silencing them and initiating a cycle of rasiRNA amplification. The sense rasiRNAs bound to Ago3 would then cleave the long, antisense transcripts produced by master regulatory loci such as *flamenco*, producing new antisense rasiRNAs that would bind to Aubergine or Piwi. If antisense transcripts from master regulatory loci are generally more abundant than sense transcripts from transposons — a reasonable assumption as transposons are normally silenced — then the pool of rasiRNAs would, as observed, be disproportionately antisense.

Of course, the model proposed by these authors[1,2] only explains how the start of each rasiRNA is defined. How the 3′ end of the rasiRNA is made remains to be discovered. And what of piRNAs in humans? piRNAs were recently discovered in immature mouse, rat and human sperm cells[10], and in zebrafish testes and ovaries[11], although they were mostly not associated with transposon sequences. What piRNAs do in mammalian sperm is unknown, but, like fly piRNAs, they derive from large genomic clusters. And like fly piRNAs[8], they do not seem to be made by Dicer. Perhaps all piRNAs are made and amplified by reciprocal cycles of Piwi-catalysed slicing of sense and antisense transcripts. Stay tuned for further detailed sequence analyses. ■

Phillip D. Zamore is in the Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.
e-mail: phillip.zamore@umassmed.edu

1. Brennecke, J. *et al. Cell* **128,** 1089–1103 (2007).
2. Gunawardane, L. S. *et al. Science* **315,** 1587–1590 (2007).
3. Fire, A. *et al. Nature* **391,** 806–811 (1998).
4. Hamilton, A. J. & Baulcombe, D. C. *Science* **286,** 950–952 (1999).
5. Ketting, R. F., Haverkamp, T. H., van Luenen, H. G. & Plasterk, R. H. *Cell* **99,** 133–141 (1999).
6. Tabara, H. *et al. Cell* **99,** 123–132 (1999).
7. Djikeng, A., Shi, H., Tschudi, C. & Ullu, E. *RNA* **7,** 1522–1530 (2001).
8. Vagin, V. V. *et al. Science* **313,** 320–324 (2006).
9. Prud'homme, N., Gans, M., Masson, M., Terzian, C. & Bucheton, A. *Genetics* **139,** 697–711 (1995).
10. Kim, V. N. *Genes Dev.* **20,** 1993–1997 (2006).
11. Houwing, S. *et al. Cell* **129,** 69–82 (2007).

NEUROBIOLOGY

Feeling right about doing right

Deborah Talmi and Chris Frith

Reason and emotion come into conflict in making all kinds of judgements. Results of work with brain-damaged patients constitute one line of evidence that the emotional component is not to be dismissed.

In resolving moral dilemmas, should emotion be our guide? This is a question prompted by various research avenues, including work described in the paper by Koenigs *et al.*[1] on page 908 of this issue.

In a typical moral dilemma, we have to choose between the lesser of two evils. Causing the death of one person is bad, but causing the death of five people is even worse. So, if you are on a runaway trolley with no other options, many people say that it is better to switch to the left fork in the track, resulting in the death of one person, than to carry on along the right fork and kill five. But what if there was no fork in the track and the only way to stop the trolley killing five people was to throw a large person, who happens to be standing next to you, under the wheels? From a utilitarian point of view the dilemma is the same: should we sacrifice one person for the sake of five? But, given this version of the dilemma, most people will choose not to throw their companion to his death. Why the difference?

There is increasing evidence that there is a strong emotional component to our moral intuitions, and that this determines, to a large degree, how we make moral judgements[2]. Thus the benefit from sacrificing a single life for the greater good must be pitted against the emotional aversion associated with the taking of life, particularly when we are face-to-face with our victim. Measurement of brain activity while people are presented with these dilemmas confirm this intuition: the moral dilemma involving throwing our companion onto the track elicits more activity in emotion-processing regions of the brain than the standard runaway-trolley problem (see ref. 3 for a review).

The implication of these ideas is that people with impaired emotional responses will have altered moral intuitions. Koenigs and his colleagues[1] have tested this hypothesis with a group of patients with damage to part of the brain called the ventral medial prefrontal cortex (VMPFC). As is typical after such damage, the autonomic nervous system in these patients showed reduced responses to emotionally charged pictures and, according to their spouses, the patients showed reduced feelings of empathy and guilt. When confronted with moral

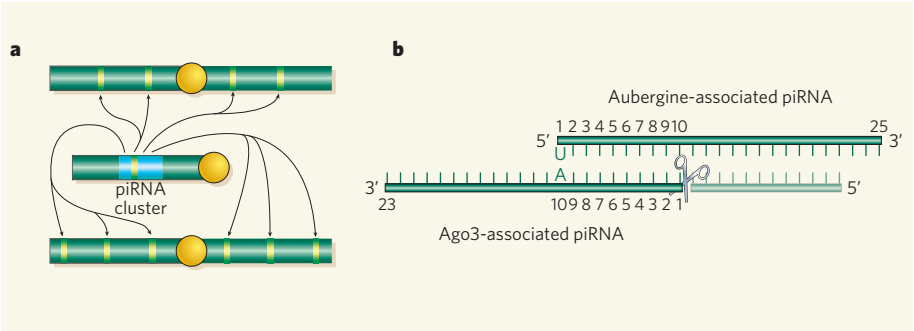


Figure 1 | piRNA-mediated silencing of transposons. **a**, A few trigger loci generate piRNAs, which target transposons (yellow) elsewhere in the genome and prevent their transcription. **b**, Complementary binding of Aubergine- and Ago3-associated piRNA sequences results in their amplification, ensuring efficient silencing of transposons.

and *ZAM* transposons and is the primary, if not the sole, source of rasiRNAs against these selfish elements.

Most rasiRNAs correspond to the antisense strand of transposons and can therefore bind to, and presumably destroy, the RNA transcripts of transposons[1,8]. These antisense rasiRNAs bind to Piwi and Aubergine. The two new papers[1,2] report that the small RNAs bound to Ago3 are nearly all of the sense orientation. Of the 353 Ago3-associated rasiRNAs identified by Siomi's group[2], the first 10 nucleotides of 16 sequences could be paired with rasiRNAs bound to Aubergine. And more than 11,200 (48%) of the Ago3-associated sense rasiRNAs identified by Hannon's group[1] could form an offset couple with at least one antisense rasiRNA.

A 10-nucleotide offset between the beginning of a small RNA guide and a second RNA molecule has a special meaning. Argonaute proteins cut target RNAs by measuring 10 nucleotides from the beginning of their RNA guide to the site of cleavage on their target. Such a pairing scheme suggests that the starting nucleotide of each antisense rasiRNA is defined by a cut that is guided by a corresponding sense rasiRNA. Reinforcing this view, nearly all the antisense rasiRNAs begin with the nucleotide U, whereas the sense rasiRNAs show no bias for beginning with U, A, C or G. Instead, the tenth nucleotide of the sense rasiRNAs was almost always A, which would allow it to pair with the first nucleotide — U — of an Aubergine- or Piwi-bound antisense rasiRNA (Fig. 1b).

Imagine, then, that a mother fly protects her offspring by providing her developing eggs with some of her Aubergine- and Piwi-bound antisense rasiRNAs. These could then generate sense rasiRNAs by cleaving the RNA transcripts of transposons, thereby simultaneously silencing them and initiating a cycle of rasiRNA amplification. The sense rasiRNAs bound to Ago3 would then cleave the long, antisense transcripts produced by master regulatory loci such as *flamenco*, producing new antisense rasiRNAs that would bind to Aubergine or Piwi. If antisense transcripts from master regulatory loci are generally more abundant than sense transcripts from transposons — a reasonable assumption as transposons are normally silenced — then the pool of rasiRNAs would, as observed, be disproportionately antisense.

Of course, the model proposed by these authors[1,2] only explains how the start of each rasiRNA is defined. How the 3′ end of the rasiRNA is made remains to be discovered. And what of piRNAs in humans? piRNAs were recently discovered in immature mouse, rat and human sperm cells[10], and in zebrafish testes and ovaries[11], although they were mostly not associated with transposon sequences. What piRNAs do in mammalian sperm is unknown, but, like fly piRNAs, they derive from large genomic clusters. And like fly piRNAs[8], they do not seem to be made by Dicer. Perhaps all piRNAs are made and amplified by reciprocal cycles of Piwi-catalysed slicing of sense and antisense transcripts. Stay tuned for further detailed sequence analyses. ■

Phillip D. Zamore is in the Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.
e-mail: phillip.zamore@umassmed.edu

1. Brennecke, J. *et al. Cell* **128,** 1089–1103 (2007).
2. Gunawardane, L. S. *et al. Science* **315,** 1587–1590 (2007).
3. Fire, A. *et al. Nature* **391,** 806–811 (1998).
4. Hamilton, A. J. & Baulcombe, D. C. *Science* **286,** 950–952 (1999).
5. Ketting, R. F., Haverkamp, T. H., van Luenen, H. G. & Plasterk, R. H. *Cell* **99,** 133–141 (1999).
6. Tabara, H. *et al. Cell* **99,** 123–132 (1999).
7. Djikeng, A., Shi, H., Tschudi, C. & Ullu, E. *RNA* **7,** 1522–1530 (2001).
8. Vagin, V. V. *et al. Science* **313,** 320–324 (2006).
9. Prud'homme, N., Gans, M., Masson, M., Terzian, C. & Bucheton, A. *Genetics* **139,** 697–711 (1995).
10. Kim, V. N. *Genes Dev.* **20,** 1993–1997 (2006).
11. Houwing, S. *et al. Cell* **129,** 69–82 (2007).

NEUROBIOLOGY

Feeling right about doing right

Deborah Talmi and Chris Frith

Reason and emotion come into conflict in making all kinds of judgements. Results of work with brain-damaged patients constitute one line of evidence that the emotional component is not to be dismissed.

In resolving moral dilemmas, should emotion be our guide? This is a question prompted by various research avenues, including work described in the paper by Koenigs *et al.*[1] on page 908 of this issue.

In a typical moral dilemma, we have to choose between the lesser of two evils. Causing the death of one person is bad, but causing the death of five people is even worse. So, if you are on a runaway trolley with no other options, many people say that it is better to switch to the left fork in the track, resulting in the death of one person, than to carry on along the right fork and kill five. But what if there was no fork in the track and the only way to stop the trolley killing five people was to throw a large person, who happens to be standing next to you, under the wheels? From a utilitarian point of view the dilemma is the same: should we sacrifice one person for the sake of five? But, given this version of the dilemma, most people will choose not to throw their companion to his death. Why the difference?

There is increasing evidence that there is a strong emotional component to our moral intuitions, and that this determines, to a large degree, how we make moral judgements[2]. Thus the benefit from sacrificing a single life for the greater good must be pitted against the emotional aversion associated with the taking of life, particularly when we are face-to-face with our victim. Measurement of brain activity while people are presented with these dilemmas confirm this intuition: the moral dilemma involving throwing our companion onto the track elicits more activity in emotion-processing regions of the brain than the standard runaway-trolley problem (see ref. 3 for a review).

The implication of these ideas is that people with impaired emotional responses will have altered moral intuitions. Koenigs and his colleagues[1] have tested this hypothesis with a group of patients with damage to part of the brain called the ventral medial prefrontal cortex (VMPFC). As is typical after such damage, the autonomic nervous system in these patients showed reduced responses to emotionally charged pictures and, according to their spouses, the patients showed reduced feelings of empathy and guilt. When confronted with moral

dilemmas, the patients with VMPFC lesions were more likely to choose the utilitarian option than were control participants and patients with lesions in other brain regions. This effect was particularly marked when there was high conflict between the utilitarian and the emotional component. The finding seems to be robust, as Ciaramelli *et al.*[4] report essentially the same results in a separate VMPFC patient group.

This result engenders a paradox. On the one hand, believing that emotion is the enemy of reason, our society still goes to great lengths to prevent emotional considerations from influencing important decisions, in particular moral decisions. Intriguingly, the result reported by Koenigs and colleagues[1] seems to show that damage to the VMPFC does not impair moral decision-making, but rather improves it through eliminating the effects of emotion. On the other hand, there is abundant evidence that decision-making in other spheres is severely impaired in these patients[5]. Indeed, it is on the basis of this latter evidence that Damasio[6] developed the 'somatic marker' theory, which affords a major role to emotions for making good decisions. Emotions clearly have a role in our moral intuitions, but do they improve moral decision-making, or impair it?

There is a similar conflict between reason and emotion when we make economic decisions. In the 'ultimatum game', player A is given $100 and can give a proportion of this to player B. Player B can accept the money, but, if B rejects it, neither A nor B get any money. Typically B will reject an offer of less than about $20. This behaviour does not accord with self-interest because B gets nothing rather than $20. It is also not utilitarian, as the two players, considered together, also get nothing rather than $100. Instead the behaviour is driven by the perceived unfairness of the offer. This is an emotional response because unfair offers elicit autonomic responses[7] and increased activity in brain regions associated with emotions[8].

Interestingly, Koenigs and Tranel[9] report that, relative to control participants, their patients with VMPFC lesions when playing the ultimatum game are influenced by emotion more strongly, in that they were more likely to reject low offers. This finding, in turn, is in line with findings that activation in the VMPFC is correlated with a reduced bias by the contextual framing of a decision[10]. Clearly, damage to this region disrupts the integration of emotion and reasoning, but it remains to be seen why this disruption takes different forms in different circumstances. Two critical differences between the paradigms Koenigs and his colleagues employed include the nature of the stimulus, which was either representational (emotional pictures, description of a dilemma)[1] or concrete (actual winning or losing a game)[9], and the procedure, which required a judgement with no direct consequences to the patient, or an active engagement[5]. Strategic planning may be needed both to stimulate an emotional response by calling forth previous experiences, and to tone down a reactive emotional response. A role for strategic planning in the emotional response that leads to the rejection of unfair offers is supported by findings[11] that disruption of function in the right prefrontal cortex reduces the effect of emotion, and biases participants towards a utilitarian response, in the ultimatum game.

Even though the precise role of VMPFC in the integration of emotion and reason remains unclear, these studies raise the perennial question as to whether emotion makes decisions better or worse. In the case of economic decisions, there is now considerable evidence that emotional rejection of unfair offers is a good strategy. Rejecting unfair offers in the ultimatum game is a form of altruistic punishment. With repeated economic encounters the use of altruistic punishment increases cooperation, resulting in greater benefits for the group[12,13]. By analogy, it may be possible to show that some moral decisions are improved when the emotional component is taken into account, rather than suppressed.

The challenge, then, is for decision-makers to cultivate an intelligent use of their emotional responses by integrating them with a reflective reasoning process, sensitive to the context and goals of the moral dilemmas they face. If decision-makers meet this challenge, they may be better able to decide when to rely upon their emotions, and when to regulate them. Indeed, such cultivation is already occurring in the legal system[14]. ■

Deborah Talmi and Chris Frith are at the Wellcome Trust Centre for Neuroimaging at University College London, 12 Queen Square, London WC1N 3BG, UK.
e-mail: cfrith@fil.ion.ucl.ac.uk

1. Koenigs, M. *et al. Nature* **446,** 908–911 (2007).
2. Haidt, J. *Psychol. Rev.* **108,** 814–834 (2001).
3. Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F. & Grafman, J. *Nature Rev. Neurosci.* **6,** 799–809 (2005).
4. Ciaramelli, E., Muccioli, M., Làdavas, E. & di Pellegrino, G. *Social Cogn. Affective Neurosci.* (in the press).
5. Bechara, A., Damasio, H. & Damasio, A. R. *Cereb. Cortex* **10,** 295–307 (2000).
6. Damasio, A. R. *Descartes' Error: Emotion, Reason, and the Human Brain* (Putnam, New York, 1994).
7. van 't Wout, M., Kahn, R. S., Sanfey, A. G. & Aleman, A. *Exp. Brain Res.* **169,** 564–568 (2006).
8. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. *Science* **300,** 1755–1758 (2003).
9. Koenigs, M. & Tranel, D. *J. Neurosci.* **27,** 951–956 (2007).
10. De Martino, B., Kumaran, D., Seymour, B. & Dolan, R. J. *Science* **313,** 684–687 (2006).
11. Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. *Science* **314,** 829–832 (2006).
12. Fehr, E. & Gächter, S. *Nature* **415,** 137–140 (2002).
13. Gurerk, O., Irlenbusch, B. & Rockenbach, B. *Science* **312,** 108–111 (2006).
14. Karstedt, S. *Theor. Criminol.* **6,** 299–317 (2002).

QUANTUM MECHANICS

To be or not to be local

Alain Aspect

The experimental violation of mathematical relations known as Bell's inequalities sounded the death-knell of Einstein's idea of 'local realism' in quantum mechanics. But which concept, locality or realism, is the problem?

The development of quantum mechanics early in the twentieth century obliged physicists to change radically the concepts they used to describe the world. The main ingredient of the first quantum revolution, wave–particle duality, has led to inventions such as the transistor and the laser that are at the root of the information society. Thanks to ideas developed by Albert Einstein and John S. Bell, another essential quantum ingredient, entanglement, is now leading us through the conceptual beginnings of a second quantum revolution — this time based on quantum information[1,2].

In contrast to wave–particle duality, which is a one-particle quantum feature, entanglement involves at least two particles. In entangled states such as those discovered by Einstein, Podolsky and Rosen (EPR)[3], quantum mechanics predicts strong correlations between measurements on two systems that have previously interacted but which are separated at the time of the measurement (Box 1). To interpret these correlations, Einstein said, one must accept the concept of local realism. This principle states that results of measurements on a system localized in space-time are fully determined by properties carried along by that system (its physical reality) and cannot be instantaneously influenced by a distant event (locality).

But after Bell's discovery that local realism entailed a limit on the correlations — a limit he expressed in his celebrated inequalities[4] — a series of ever more ideal experiments (ref. 5 and references therein) has led us to abandon the concept. It is then natural to raise the question of whether one should drop locality — which equates to the impossibility of any influence travelling faster than light — or rather drop the notion of physical reality.

There is no logical answer to that question: one can choose to abandon either concept, or even both. Tony Leggett has explored[6] one of these possibilities by considering a particular class of physically plausible theories that abandon locality, but maintain realism. He found these theories to be incompatible with quantum mechanics, and expressed the disagreement by new inequalities[6]. As there was no experimental

dilemmas, the patients with VMPFC lesions were more likely to choose the utilitarian option than were control participants and patients with lesions in other brain regions. This effect was particularly marked when there was high conflict between the utilitarian and the emotional component. The finding seems to be robust, as Ciaramelli *et al.*[4] report essentially the same results in a separate VMPFC patient group.

This result engenders a paradox. On the one hand, believing that emotion is the enemy of reason, our society still goes to great lengths to prevent emotional considerations from influencing important decisions, in particular moral decisions. Intriguingly, the result reported by Koenigs and colleagues[1] seems to show that damage to the VMPFC does not impair moral decision-making, but rather improves it through eliminating the effects of emotion. On the other hand, there is abundant evidence that decision-making in other spheres is severely impaired in these patients[5]. Indeed, it is on the basis of this latter evidence that Damasio[6] developed the 'somatic marker' theory, which affords a major role to emotions for making good decisions. Emotions clearly have a role in our moral intuitions, but do they improve moral decision-making, or impair it?

There is a similar conflict between reason and emotion when we make economic decisions. In the 'ultimatum game', player A is given $100 and can give a proportion of this to player B. Player B can accept the money, but, if B rejects it, neither A nor B get any money. Typically B will reject an offer of less than about $20. This behaviour does not accord with self-interest because B gets nothing rather than $20. It is also not utilitarian, as the two players, considered together, also get nothing rather than $100. Instead the behaviour is driven by the perceived unfairness of the offer. This is an emotional response because unfair offers elicit autonomic responses[7] and increased activity in brain regions associated with emotions[8].

Interestingly, Koenigs and Tranel[9] report that, relative to control participants, their patients with VMPFC lesions when playing the ultimatum game are influenced by emotion more strongly, in that they were more likely to reject low offers. This finding, in turn, is in line with findings that activation in the VMPFC is correlated with a reduced bias by the contextual framing of a decision[10]. Clearly, damage to this region disrupts the integration of emotion and reasoning, but it remains to be seen why this disruption takes different forms in different circumstances. Two critical differences between the paradigms Koenigs and his colleagues employed include the nature of the stimulus, which was either representational (emotional pictures, description of a dilemma)[1] or concrete (actual winning or losing a game)[9], and the procedure, which required a judgement with no direct consequences to the patient, or an active engagement[5]. Strategic planning may be needed both to stimulate an emotional response by calling forth previous experiences, and to tone down a reactive emotional response. A role for strategic planning in the emotional response that leads to the rejection of unfair offers is supported by findings[11] that disruption of function in the right prefrontal cortex reduces the effect of emotion, and biases participants towards a utilitarian response, in the ultimatum game.

Even though the precise role of VMPFC in the integration of emotion and reason remains unclear, these studies raise the perennial question as to whether emotion makes decisions better or worse. In the case of economic decisions, there is now considerable evidence that emotional rejection of unfair offers is a good strategy. Rejecting unfair offers in the ultimatum game is a form of altruistic punishment. With repeated economic encounters the use of altruistic punishment increases cooperation, resulting in greater benefits for the group[12,13]. By analogy, it may be possible to show that some moral decisions are improved when the emotional component is taken into account, rather than suppressed.

The challenge, then, is for decision-makers to cultivate an intelligent use of their emotional responses by integrating them with a reflective reasoning process, sensitive to the context and goals of the moral dilemmas they face. If decision-makers meet this challenge, they may be better able to decide when to rely upon their emotions, and when to regulate them. Indeed, such cultivation is already occurring in the legal system[14]. ■

Deborah Talmi and Chris Frith are at the Wellcome Trust Centre for Neuroimaging at University College London, 12 Queen Square, London WC1N 3BG, UK.
e-mail: cfrith@fil.ion.ucl.ac.uk

1. Koenigs, M. *et al. Nature* **446,** 908–911 (2007).
2. Haidt, J. *Psychol. Rev.* **108,** 814–834 (2001).
3. Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F. & Grafman, J. *Nature Rev. Neurosci.* **6,** 799–809 (2005).
4. Ciaramelli, E., Muccioli, M., Làdavas, E. & di Pellegrino, G. *Social Cogn. Affective Neurosci.* (in the press).
5. Bechara, A., Damasio, H. & Damasio, A. R. *Cereb. Cortex* **10,** 295–307 (2000).
6. Damasio, A. R. *Descartes' Error: Emotion, Reason, and the Human Brain* (Putnam, New York, 1994).
7. van 't Wout, M., Kahn, R. S., Sanfey, A. G. & Aleman, A. *Exp. Brain Res.* **169,** 564–568 (2006).
8. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. *Science* **300,** 1755–1758 (2003).
9. Koenigs, M. & Tranel, D. *J. Neurosci.* **27,** 951–956 (2007).
10. De Martino, B., Kumaran, D., Seymour, B. & Dolan, R. J. *Science* **313,** 684–687 (2006).
11. Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. *Science* **314,** 829–832 (2006).
12. Fehr, E. & Gächter, S. *Nature* **415,** 137–140 (2002).
13. Gurerk, O., Irlenbusch, B. & Rockenbach, B. *Science* **312,** 108–111 (2006).
14. Karstedt, S. *Theor. Criminol.* **6,** 299–317 (2002).

QUANTUM MECHANICS

To be or not to be local

Alain Aspect

The experimental violation of mathematical relations known as Bell's inequalities sounded the death-knell of Einstein's idea of 'local realism' in quantum mechanics. But which concept, locality or realism, is the problem?

The development of quantum mechanics early in the twentieth century obliged physicists to change radically the concepts they used to describe the world. The main ingredient of the first quantum revolution, wave–particle duality, has led to inventions such as the transistor and the laser that are at the root of the information society. Thanks to ideas developed by Albert Einstein and John S. Bell, another essential quantum ingredient, entanglement, is now leading us through the conceptual beginnings of a second quantum revolution — this time based on quantum information[1,2].

In contrast to wave–particle duality, which is a one-particle quantum feature, entanglement involves at least two particles. In entangled states such as those discovered by Einstein, Podolsky and Rosen (EPR)[3], quantum mechanics predicts strong correlations between measurements on two systems that have previously interacted but which are separated at the time of the measurement (Box 1). To interpret these correlations, Einstein said, one must accept the concept of local realism. This principle states that results of measurements on a system localized in space-time are fully determined by properties carried along by that system (its physical reality) and cannot be instantaneously influenced by a distant event (locality).

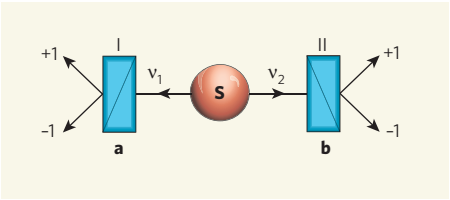
But after Bell's discovery that local realism entailed a limit on the correlations — a limit he expressed in his celebrated inequalities[4] — a series of ever more ideal experiments (ref. 5 and references therein) has led us to abandon the concept. It is then natural to raise the question of whether one should drop locality — which equates to the impossibility of any influence travelling faster than light — or rather drop the notion of physical reality.

There is no logical answer to that question: one can choose to abandon either concept, or even both. Tony Leggett has explored[6] one of these possibilities by considering a particular class of physically plausible theories that abandon locality, but maintain realism. He found these theories to be incompatible with quantum mechanics, and expressed the disagreement by new inequalities[6]. As there was no experimental

Box 1 | Thought made reality

As discovered by Albert Einstein, Boris Podolsky and Nathan Rosen[3], quantum mechanics predicts strong correlations between measurements on two particles in an entangled state. It is tempting to interpret these correlations as the result of shared properties determined at the time of their initial interaction and carried along by each particle. By analogy, similar sets of chromosomes in siblings allow one to understand correlations in their eye colour or other features.

Theories completed in such a way implement a view of the physical world called local realism, because individual physical properties are attributed to each of the separated partners. This type of interpretation was favoured by Einstein, but strongly opposed by another great early quantum physicist, Niels Bohr. For decades, however, the opposition between Einstein and Bohr seemed to be a mere epistemological debate, without any consequences for the predictions of the theory.



John Bell's formulation of his celebrated inequalities[4], which fix a limit to the correlations predicted by local realistic theories, made it possible to settle the debate by performing an experiment to test the inequalities. For a well-designed experiment[11], quantum mechanics predicts a violation of Bell's inequalities. There were thus two possibilities, both interesting: either the experimental results would obey Bell's inequalities, and thus exhibit a failure of quantum mechanics, or they would violate Bell's inequalities, and force us to renounce Einstein's local realist world view.

Starting with the pioneering work of John Clauser[12], a series of more and more refined experiments has brought overwhelming evidence that the actual degree of correlation found experimentally indeed violates Bell's inequalities[5]. In the basic experimental realization pictured here, a pair of photons, ν_1 and ν_2, is produced at source S in an entangled polarization state. Each photon is submitted to a measurement by a linear polarizer (I and II, in orientations **a** and **b**), giving a result +1 or −1. Quantum mechanics predicts that individual outcomes happen at random with equal probabilities at each polarizer, but that outcomes of both sides will be strongly correlated — as indeed they are.

The role of locality in these experiments has been underlined by changing the setting of the polarizers while the photons are in flight, and by making sure that the two measurements cannot influence each other according to relativistic causality[13–15]. The clear violation of Bell's inequalities leads to the conclusive rejection of theories that are simultaneously realistic and local. **A. A.**

result available to test Leggett's inequalities, a new type of measurement was necessary.

In experiments detailed on page 871 of this issue[7], Gröblacher *et al.* have carried out such measurements. They modified an experiment previously used to test Bell's inequalities through measurements on two photons (Box 1) by changing the original linear polarization measurement into an elliptical polarization measurement. This is readily done by inserting a quarter-wave plate in front of a standard (linear) polarizer. From an experimental point of view, however, the new test is more demanding, and requires almost ideal optical elements and a high signal-to-noise ratio. Thanks to their high-efficiency source of entangled photons, the authors meet these requirements, and find a significant violation of the generalized Leggett's inequalities that they have established. Following Leggett, they conclude by questioning realism rather than locality — at variance with the often-heard statement that "quantum mechanics is non-local".

Interesting as this conclusion is, it remains a matter of personal preference, not of logical deduction. The violation of Bell's inequalities implied that realism and locality are not simultaneously tenable. Violation of Leggett's inequalities implies only that realism and a certain type of non-locality are incompatible: there are other types of non-local models that are not addressed by either Leggett's inequalities or the experiment.

Consider, for instance, the experiment pictured in Box 1, and assume that a measurement is performed first on photon ν_1. This measurement gives either a result of +1 or −1; immediately after one of the two results is obtained, the quantum description of ν_2, which had not been favouring any precise polarization before the measurement on ν_1, collapses into a state of polarization identical to the one found for photon ν_1, from which one can readily derive the usual quantum-mechanical EPR correlations. If we take this description — based on standard quantum-mechanical calculations — as a model, it cannot be rejected by any experiment that is in agreement with quantum mechanics, including the more complex elliptical polarization measurements performed by Gröblacher *et al.*[7].

This model is clearly non-local in the relativistic sense, as we must invoke a particular frame of reference to give a sense to the statements that measurement on ν_1 happens first, and that its result immediately affects the state of ν_2. Can we say that the model is realist? In a sense yes, as we can qualify the individual polarization of each photon at each step. I must admit, however, that I am not comfortable with the notion of a physical reality that is instantaneously modified by something happening far away — not to speak of the problems related to breaking down explicit relativistic covariance[8]. In many of Einstein's writings[9], the notion of a physical reality describing completely the state of a system localized in space-time is clearly linked to the relativistic impossibility that this physical reality be instantaneously modified by a faraway event. To quote Einstein[10], the necessity of completing quantum mechanics in a local–realist way could be escaped "only by either assuming that the measurement of S_1 (telepathically) changes the real situation of S_2 or by denying independent real situations to things which are spatially separated from each other. Both alternatives appear to me entirely unacceptable."

Nobody knows what Einstein would have thought had he known of the violation of Bell's and Leggett's inequalities. I tend to accept the kind of non-local image sketched above as useful to stimulate my imagination, although I am aware that it implies renouncing the kind of realism I would have liked. The conclusion one draws is more a question of taste than logic, and one can argue that such a discussion is irrelevant to science. But I rather share the view that such debates, and accompanying experiments such as those of Gröblacher *et al.*[7], allow us to look deeper into the great mysteries of quantum mechanics. That, it is to be hoped, can contribute to transforming the second quantum revolution from the present stage of basic research to a fully fledged technological revolution. ■

Alain Aspect is in the Laboratoire Charles Fabry de l'Institut d'Optique, Campus Polytechnique, RD 128, F-91127 Palaiseau, France.
e-mail: alain.aspect@institutoptique.fr

1. Dowling, J. & Milburn, G. *Phil. Trans. R. Soc. Lond. A* **361,** 1655–1674 (2003).
2. Aspect, A. in Bell, J. *Speakable and Unspeakable in Quantum Mechanics*, xvii–xxxix (Cambridge Univ. Press, 2004).
3. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **47,** 777–780 (1935).
4. Bell, J. *Physics* **1,** 195–200 (1964); reprinted in Bell, J. *Speakable and Unspeakable in Quantum Mechanics* 14–21 (Cambridge Univ. Press, 2004).
5. Aspect, A. *Nature* **398,** 189–190 (1999).
6. Leggett, A. J. *Found. Phys.* **33,** 1469–1493 (2003).
7. Gröblacher, S. *et al. Nature* **446,** 871–875 (2007).
8. Peres, A. *Phys. Rev. A* **61,** 022117 (2000).
9. Born, M. *The Born–Einstein Letters* 1916–1955 (Macmillan, London, 1971).
10. Schilpp, P. A. (ed.) *A. Einstein: Philosopher Scientist,* 85 (Open Court and Cambridge Univ. Press, 1949).
11. Clauser, J. F., Horne, M. A, Shimony, R. A. & Holt, R. A. *Phys. Rev. Lett.* **23,** 880–884 (1969).
12. Freedman, S. J. & Clauser, J. F. *Phys. Rev. Lett.* **28,** 938–941 (1972).
13. Aspect, A., Dalibard, J. & Roger, G. *Phys. Rev. Lett.* **49,** 1804–1807 (1982).
14. Weihs, G., Jennewein, T., Simon, C., Weinfurter, H. & Zeilinger, A. *Phys. Rev. Lett.* **81,** 5039–5042 (1998).
15. Tittel, W., Brendel, J., Zbinden, H. & Gisin, N. *Phys. Rev. Lett.* **81,** 3563–3566 (1998).

CELL BIOLOGY

The checkpoint brake relieved

Jan-Michael Peters

When a cell divides, each daughter cell inherits a complete set of chromosomes. A sophisticated inhibitory mechanism delays chromosome segregation and cell division until everything is in its place.

It is equally important for a cell to do things at the right times as it is for an organism. This is particularly true during the final phase of cell division — mitosis — when chromosomes segregate. Chromosome segregation must not be initiated too early, because otherwise the daughter cells will carry an abnormal number of chromosomes and, in rare cases, tumour cells or diseases such as Down's syndrome might develop. So, a special surveillance system known as the spindle-assembly checkpoint inhibits chromosome segregation until this process can be carried out properly. How this system works is one of the big mysteries of the cell cycle. On pages 876 and 921 of this issue, two studies from the laboratories of Elledge[1] and Kirschner[2] propose an intriguing model for how the activity of the spindle-assembly checkpoint may be maintained, and eventually relieved, when the cell is ready to separate its chromosomes.

Long before a cell divides, it generates an extra copy of its chromosomes by DNA replication, so that every chromosome contains two identical copies of DNA, packaged into two halves of the chromosome — the sister chromatids. During mitosis, sister chromatids are separated and are transported towards opposite poles of the cell so that, after division, each daughter cell carries an identical set of sister chromatids. For this to occur, it is essential that the two sister chromatids become attached to the opposite halves of the mitotic spindle before they are separated from each other. The spindle-assembly checkpoint ensures this by 'sensing' the presence of chromosomes whose two kinetochores (spindle-attachment sites) are not properly attached to the spindle[3]. As long as a dividing cell contains such unattached kinetochores, the spindle-assembly checkpoint is active and prevents the separation of sister chromatids[4] (Fig. 1a).

The spindle-assembly checkpoint halts cell division by inhibiting the anaphase-promoting complex/cyclosome (APC/C)[5–7] — an enzyme that adds chains of the small protein ubiquitin to specific substrate proteins, thereby targeting them for destruction[8]. During cell division, these reactions are mediated by APC/C and two other proteins — Cdc20, which helps substrate recruitment to APC/C, and UbcH10, which is a ubiquitin-conjugating enzyme that transfers ubiquitin residues onto the substrate proteins.

These ubiquitinating reactions are inhibited when unattached kinetochores keep the spindle-assembly checkpoint active. The kinetochores promote the association of Cdc20 with at least two checkpoint proteins known as Mad2 and BubR1, which prevent Cdc20 from activating the APC/C[9]. But this inhibition can only persist until all the chromosomes are attached to both poles of the spindle, because soon after this the APC/C becomes active and sister chromatids begin to separate.

How, then, is Cdc20 relieved from its inhibition? It has often been assumed that the Cdc20–Mad2–BubR1 complex is simply short-lived, and that the inactivation of the spindle-assembly checkpoint might therefore passively release Cdc20 from its inhibitors. However, structural studies of the Mad2–Cdc20 interaction have indicated that Mad2, at least, holds on to Cdc20 very tightly, like the way that a safety belt secures a passenger in a car seat[10,11]. Thus it is plausible that the release of Cdc20 from Mad2 is an active process.

This is exactly what the two papers from the Kirschner and Elledge groups propose. More surprisingly, these authors provide evidence to suggest that the APC/C itself might be the molecule that liberates Cdc20 from inhibition by Mad2 and BubR1 (Fig. 1b).

The Kirschner[2] lab came to this conclusion by studying extracts from cells with an active spindle-assembly checkpoint, which, not surprisingly, contained little or no APC/C activity. However, when UbcH10 was added to these extracts, APC/C became active, coincident with the ubiquitination of Cdc20 and the dissociation of Mad2 and BubR1 from Cdc20. Ubiquitination often targets proteins for destruction by the 26S proteasome enzyme complex, but Kirschner's team found that proteasome inhibition did not prevent the activation of APC/C by UbcH10. On the basis of these and other observations, the authors propose that the

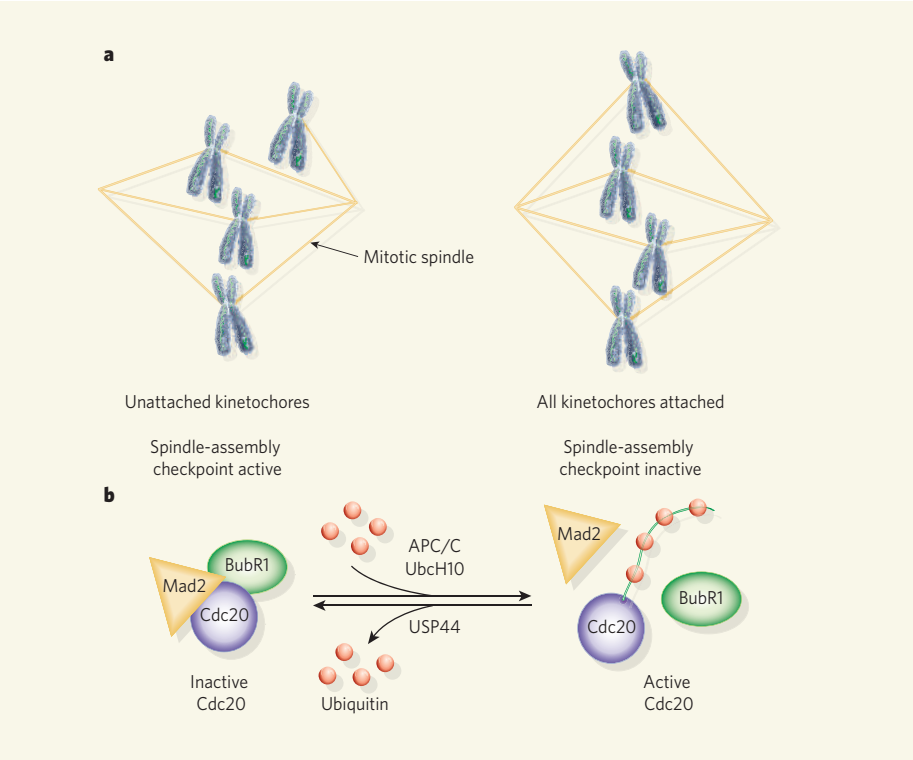


Figure 1 | Regulation of cell division by the spindle-assembly checkpoint. a, Before a cell divides, the sister chromatids from all of its chromosomes have to attach to the two poles of the mitotic spindle. As long as chromosomes are present that do not obey this rule, a cellular surveillance system known as the spindle-assembly checkpoint is active. This inhibits the continuation of cell division and therefore prevents an abnormal number of sister chromatids reaching each daughter cell. **b**, At the molecular level, the spindle-assembly checkpoint proteins Mad2 and BubR1 associate with Cdc20 and inhibit its ability to activate the ubiquitin ligase APC/C, which is required for the progression of cell division. According to a model proposed by Elledge *et al.*[1] and Kirschner *et al.*[2], APC/C and the ubiquitin-conjugating enzyme UbcH10 can overcome this inhibition by ubiquitinating Cdc20, which causes Mad2 and BubR1 to become dissociated from it. Cdc20 ubiquitination is, in turn, reversed by the de-ubiquitinating enzyme USP44. A balance between ubiquitination and de-ubiquitination reactions may therefore control the activity of Cdc20. Once all chromosomes have been properly attached to the mitotic spindle, this balance would tip towards ubiquitination of Cdc20, with Cdc20 and the APC/C subsequently being able to initiate the separation of sister chromatids.

SOCIAL SCIENCE

The urban organism

Visitors to the area around *Nature*'s London offices will be familiar with the scene: unending traffic and noise; the hurly-burly of the Underground; streets, concourses and platforms filled with people intent only on reaching their destination quickly. It's received wisdom that the bigger a city is, the faster life moves; Luis Bettencourt and colleagues supply some empirical evidence to back up that perception (*Proc. Natl Acad. Sci. USA* **104,** 7301–7306; 2007).

The authors begin by examining how different indicators of cities' activity and infrastructure scale with their size. They use various sets of data from the United States, China and Germany, and characterize the scalings as power laws of the form (population)^n. They find that indicators of economic activity — from personal income, to patent registrations, to total electricity consumption — vary with population with values of n in the range 1.1–1.3, regardless of where the data were collected. In other words, cities the world over become more hyperactive the larger they get. Perhaps as a corollary to that excess, the prevalence of crime and sexually transmitted disease grows similarly quickly.

Infrastructure indicators such as the lengths of the road and electricity networks, by contrast, scale to around (population)$^{0.8}$. The larger the metropolis, the less of these things each citizen has at their disposal. Thus it seems that cities fulfill two basic needs of modern human society: they facilitate the exchange of ideas and, by extension, wealth creation; and they achieve economies of scale in the supply of a population's needs.

To look at how these very different dynamics affect city expansion over time, Bettencourt *et al.* construct a general equation that models the cost on resources of sustaining and increasing a population. Unsurprisingly, growth driven by the demands of efficiency, $n < 1$, stagnates after time: economies of scale eventually hit a bottom line.

City growth driven by wealth creation ($n > 1$), on the other hand, rapidly becomes hyperexponential. The only way to avoid collapse as a population outstrips the finite resources available to it is through constant waves of innovation. These effectively re-engineer the initial conditions of growth. But the greater the absolute population, the smaller the relative return on each such investment — so new ideas must come ever faster.

The city dweller looking for a quiet life is thus hit with a double whammy: the bigger the city, the faster life is; but the rate at which life gets faster must itself accelerate to maintain the city as a going concern.

In biological organisms, the authors note, the situation is completely different. Larger organisms have greater economies of scale, and slower-paced lives. Metabolic rates, for example, increase with (body mass)$^{0.75}$. With the city, it seems, mankind has created an organism operating beyond the bounds of what is natural. ■

Richard Webb

addition of a ubiquitin chain to Cdc20 by APC/C and UbcH10 does not necessarily involve protein degradation, but leads to the dissociation of Mad2 and BubR1 from Cdc20. One implication of this model is that APC/C would constantly antagonize its inhibition by the spindle-assembly checkpoint. If so, how could Mad2 and BubR1 ever inhibit the APC/C in cells with an active checkpoint?

A possible answer comes from a study carried out by Elledge's team[1]. In a search for proteins that are required for the activity of the spindle-assembly checkpoint, these authors identified a de-ubiquitinating enzyme known as USP44. Enzymes of this type disassemble ubiquitin chains by cleaving the bonds that connect the ubiquitin residues in the chain. Interestingly, USP44 differs from other known spindle-assembly checkpoint proteins in that it is not required to recruit Mad2 to unattached kinetochores, where Mad2 is believed to form complexes with Cdc20 and BubR1. So how else could USP44 function at the checkpoint? It turns out that, *in vitro*, USP44 can inhibit the ability of UbcH10 to activate checkpoint-inhibited APC/C, leading Elledge and colleagues to propose that USP44 might stabilize Cdc20–Mad2–BubR1 complexes by destroying the ubiquitin chains that APC/C adds to Cdc20 (Fig. 1b). Consistent with this argument, depletion of USP44 prematurely inactivates the spindle-assembly checkpoint in mitotic cells and leads to defects in chromosome segregation.

The model proposing that the stability of Cdc20–Mad2–BubR1 complexes is controlled by a fine balance between ubiquitination, mediated by the APC/C, and de-ubiquitination, catalysed by USP44, makes a number of predictions. Testing these will be an essential goal for the future. For example, could a mutant of Cdc20 be created that couldn't be ubiquitinated but would otherwise be functional? If so, such a mutant would be predicted to assemble into unusually stable checkpoint complexes from which Mad2 and BubR1 could not easily dissociate.

These studies[1,2] also raise a number of other questions. Is de-ubiquitinating Cdc20 the main role of USP44 in maintaining the spindle-assembly checkpoint, or does it also antagonize APC/C more directly by disassembling ubiquitin chains on its other protein substrates such as securin and cyclin B? How does ubiquitination dissociate Mad2 and BubR1 from Cdc20 — by inducing conformational changes in these proteins, or by recruiting enzymes (such as the p97/Cdc48–ATPase) that would catalyse the dissociation process? Finally, how is the balance between de-ubiquitination and ubiquitination reactions tipped once all chromosomes have become attached to both poles of the mitotic spindle? Elledge *et al.* found that USP44 itself is degraded at the end of cell division. Could this be the primary switch for checkpoint inactivation, or is it merely a consequence of APC/C activation once the checkpoint has been silenced? Answering these questions will keep researchers busy for some time to come. ■

Jan-Michael Peters is at the Research Institute of Molecular Pathology (IMP), Dr. Bohr-Gasse 7, A-1030 Vienna, Austria.
e-mail: peters@imp.univie.ac.at

1. Stegmeier, F. *et al. Nature* **446,** 876–880 (2007).
2. Reddy, S. K., Rape, M., Marganski, W. A. & Kirschner, M. W. *Nature* **446,** 921–925 (2007).
3. Rieder, C. L., Schultz, A., Cole, R. & Sluder, G. *J. Cell Biol.* **127,** 1301–1310 (1994).
4. Nasmyth, K. *Cell* **120,** 739–746 (2005).
5. Li, Y., Gorbea, C., Mahaffey, D., Rechsteiner, M. & Benezra, R. *Proc. Natl Acad. Sci. USA* **94,** 12431–12436 (1997).
6. Hwang, L. H. *et al. Science* **279,** 1041–1044 (1998).
7. Kim, S. H., Lin, D. P., Matsumoto, S., Kitazono, A. & Matsumoto, T. *Science* **279,** 1045–1047 (1998).
8. Peters, J.-M. *Nature Rev. Mol. Cell Biol.* **7,** 644–656 (2006).
9. Sudakin, V., Chan, G. K. & Yen, T. J. *J. Cell Biol.* **154,** 925–936 (2001).
10. Luo, X., Tang, Z., Rizo, J. & Yu, H. *Mol. Cell* **9,** 59–71 (2002).
11. Sironi, L. *et al. EMBO J.* **21,** 2496–2506 (2002).

Correction
In the News & Views article "Cell biology: Lost in mitotic translation" by Anthony Wynshaw-Boris (*Nature* **446,** 274–275; 2007), statements in the text and the caption to Figure 1, and part **b** of Figure 1, imply that eIF4B binds directly to the 5′ cap of mRNA. Rather, eIF4B facilitates the ATP-dependent helicase activity of eIF4A to promote the ribosome recruitment necessary for cap-dependent translation.

SOCIAL SCIENCE

The urban organism

Visitors to the area around *Nature*'s London offices will be familiar with the scene: unending traffic and noise; the hurly-burly of the Underground; streets, concourses and platforms filled with people intent only on reaching their destination quickly. It's received wisdom that the bigger a city is, the faster life moves; Luis Bettencourt and colleagues supply some empirical evidence to back up that perception (*Proc. Natl Acad. Sci. USA* **104,** 7301–7306; 2007).

The authors begin by examining how different indicators of cities' activity and infrastructure scale with their size. They use various sets of data from the United States, China and Germany, and characterize the scalings as power laws of the form (population)^n. They find that indicators of economic activity — from personal income, to patent registrations, to total electricity consumption — vary with population with values of n in the range 1.1–1.3, regardless of where the data were collected. In other words, cities the world over become more hyperactive the larger they get. Perhaps as a corollary to that excess, the prevalence of crime and sexually transmitted disease grows similarly quickly.

Infrastructure indicators such as the lengths of the road and electricity networks, by contrast, scale to around (population)$^{0.8}$. The larger the metropolis, the less of these things each citizen has at their disposal. Thus it seems that cities fulfill two basic needs of modern human society: they facilitate the exchange of ideas and, by extension, wealth creation; and they achieve economies of scale in the supply of a population's needs.

To look at how these very different dynamics affect city expansion over time, Bettencourt *et al.* construct a general equation that models the cost on resources of sustaining and increasing a population. Unsurprisingly, growth driven by the demands of efficiency, $n < 1$, stagnates after time: economies of scale eventually hit a bottom line.

City growth driven by wealth creation ($n > 1$), on the other hand, rapidly becomes hyperexponential. The only way to avoid collapse as a population outstrips the finite resources available to it is through constant waves of innovation. These effectively re-engineer the initial conditions of growth. But the greater the absolute population, the smaller the relative return on each such investment — so new ideas must come ever faster.

The city dweller looking for a quiet life is thus hit with a double whammy: the bigger the city, the faster life is; but the rate at which life gets faster must itself accelerate to maintain the city as a going concern.

In biological organisms, the authors note, the situation is completely different. Larger organisms have greater economies of scale, and slower-paced lives. Metabolic rates, for example, increase with (body mass)$^{0.75}$. With the city, it seems, mankind has created an organism operating beyond the bounds of what is natural. ■

Richard Webb

addition of a ubiquitin chain to Cdc20 by APC/C and UbcH10 does not necessarily involve protein degradation, but leads to the dissociation of Mad2 and BubR1 from Cdc20. One implication of this model is that APC/C would constantly antagonize its inhibition by the spindle-assembly checkpoint. If so, how could Mad2 and BubR1 ever inhibit the APC/C in cells with an active checkpoint?

A possible answer comes from a study carried out by Elledge's team[1]. In a search for proteins that are required for the activity of the spindle-assembly checkpoint, these authors identified a de-ubiquitinating enzyme known as USP44. Enzymes of this type disassemble ubiquitin chains by cleaving the bonds that connect the ubiquitin residues in the chain. Interestingly, USP44 differs from other known spindle-assembly checkpoint proteins in that it is not required to recruit Mad2 to unattached kinetochores, where Mad2 is believed to form complexes with Cdc20 and BubR1. So how else could USP44 function at the checkpoint? It turns out that, *in vitro*, USP44 can inhibit the ability of UbcH10 to activate checkpoint-inhibited APC/C, leading Elledge and colleagues to propose that USP44 might stabilize Cdc20–Mad2–BubR1 complexes by destroying the ubiquitin chains that APC/C adds to Cdc20 (Fig. 1b). Consistent with this argument, depletion of USP44 prematurely inactivates the spindle-assembly checkpoint in mitotic cells and leads to defects in chromosome segregation.

The model proposing that the stability of Cdc20–Mad2–BubR1 complexes is controlled by a fine balance between ubiquitination, mediated by the APC/C, and de-ubiquitination, catalysed by USP44, makes a number of predictions. Testing these will be an essential goal for the future. For example, could a mutant of Cdc20 be created that couldn't be ubiquitinated but would otherwise be functional? If so, such a mutant would be predicted to assemble into unusually stable checkpoint complexes from which Mad2 and BubR1 could not easily dissociate.

These studies[1,2] also raise a number of other questions. Is de-ubiquitinating Cdc20 the main role of USP44 in maintaining the spindle-assembly checkpoint, or does it also antagonize APC/C more directly by disassembling ubiquitin chains on its other protein substrates such as securin and cyclin B? How does ubiquitination dissociate Mad2 and BubR1 from Cdc20 — by inducing conformational changes in these proteins, or by recruiting enzymes (such as the p97/Cdc48–ATPase) that would catalyse the dissociation process? Finally, how is the balance between de-ubiquitination and ubiquitination reactions tipped once all chromosomes have become attached to both poles of the mitotic spindle? Elledge *et al.* found that USP44 itself is degraded at the end of cell division. Could this be the primary switch for checkpoint inactivation, or is it merely a consequence of APC/C activation once the checkpoint has been silenced? Answering these questions will keep researchers busy for some time to come. ■

Jan-Michael Peters is at the Research Institute of Molecular Pathology (IMP), Dr. Bohr-Gasse 7, A-1030 Vienna, Austria.
e-mail: peters@imp.univie.ac.at

1. Stegmeier, F. *et al. Nature* **446,** 876–880 (2007).
2. Reddy, S. K., Rape, M., Marganski, W. A. & Kirschner, M. W. *Nature* **446,** 921–925 (2007).
3. Rieder, C. L., Schultz, A., Cole, R. & Sluder, G. *J. Cell Biol.* **127,** 1301–1310 (1994).
4. Nasmyth, K. *Cell* **120,** 739–746 (2005).
5. Li, Y., Gorbea, C., Mahaffey, D., Rechsteiner, M. & Benezra, R. *Proc. Natl Acad. Sci. USA* **94,** 12431–12436 (1997).
6. Hwang, L. H. *et al. Science* **279,** 1041–1044 (1998).
7. Kim, S. H., Lin, D. P., Matsumoto, S., Kitazono, A. & Matsumoto, T. *Science* **279,** 1045–1047 (1998).
8. Peters, J.-M. *Nature Rev. Mol. Cell Biol.* **7,** 644–656 (2006).
9. Sudakin, V., Chan, G. K. & Yen, T. J. *J. Cell Biol.* **154,** 925–936 (2001).
10. Luo, X., Tang, Z., Rizo, J. & Yu, H. *Mol. Cell* **9,** 59–71 (2002).
11. Sironi, L. *et al. EMBO J.* **21,** 2496–2506 (2002).

Correction
In the News & Views article "Cell biology: Lost in mitotic translation" by Anthony Wynshaw-Boris (*Nature* **446,** 274–275; 2007), statements in the text and the caption to Figure 1, and part **b** of Figure 1, imply that eIF4B binds directly to the 5′ cap of mRNA. Rather, eIF4B facilitates the ATP-dependent helicase activity of eIF4A to promote the ribosome recruitment necessary for cap-dependent translation.

SOCIAL SCIENCE

The urban organism

Visitors to the area around *Nature*'s London offices will be familiar with the scene: unending traffic and noise; the hurly-burly of the Underground; streets, concourses and platforms filled with people intent only on reaching their destination quickly. It's received wisdom that the bigger a city is, the faster life moves; Luis Bettencourt and colleagues supply some empirical evidence to back up that perception (*Proc. Natl Acad. Sci. USA* **104,** 7301–7306; 2007).

The authors begin by examining how different indicators of cities' activity and infrastructure scale with their size. They use various sets of data from the United States, China and Germany, and characterize the scalings as power laws of the form (population)^n. They find that indicators of economic activity — from personal income, to patent registrations, to total electricity consumption — vary with population with values of n in the range 1.1–1.3, regardless of where the data were collected. In other words, cities the world over become more hyperactive the larger they get. Perhaps as a corollary to that excess, the prevalence of crime and sexually transmitted disease grows similarly quickly.

Infrastructure indicators such as the lengths of the road and electricity networks, by contrast, scale to around (population)$^{0.8}$. The larger the metropolis, the less of these things each citizen has at their disposal. Thus it seems that cities fulfill two basic needs of modern human society: they facilitate the exchange of ideas and, by extension, wealth creation; and they achieve economies of scale in the supply of a population's needs.

To look at how these very different dynamics affect city expansion over time, Bettencourt *et al.* construct a general equation that models the cost on resources of sustaining and increasing a population. Unsurprisingly, growth driven by the demands of efficiency, $n < 1$, stagnates after time: economies of scale eventually hit a bottom line.

City growth driven by wealth creation ($n > 1$), on the other hand, rapidly becomes hyperexponential. The only way to avoid collapse as a population outstrips the finite resources available to it is through constant waves of innovation. These effectively re-engineer the initial conditions of growth. But the greater the absolute population, the smaller the relative return on each such investment — so new ideas must come ever faster.

The city dweller looking for a quiet life is thus hit with a double whammy: the bigger the city, the faster life is; but the rate at which life gets faster must itself accelerate to maintain the city as a going concern.

In biological organisms, the authors note, the situation is completely different. Larger organisms have greater economies of scale, and slower-paced lives. Metabolic rates, for example, increase with (body mass)$^{0.75}$. With the city, it seems, mankind has created an organism operating beyond the bounds of what is natural. ■

Richard Webb

addition of a ubiquitin chain to Cdc20 by APC/C and UbcH10 does not necessarily involve protein degradation, but leads to the dissociation of Mad2 and BubR1 from Cdc20. One implication of this model is that APC/C would constantly antagonize its inhibition by the spindle-assembly checkpoint. If so, how could Mad2 and BubR1 ever inhibit the APC/C in cells with an active checkpoint?

A possible answer comes from a study carried out by Elledge's team[1]. In a search for proteins that are required for the activity of the spindle-assembly checkpoint, these authors identified a de-ubiquitinating enzyme known as USP44. Enzymes of this type disassemble ubiquitin chains by cleaving the bonds that connect the ubiquitin residues in the chain. Interestingly, USP44 differs from other known spindle-assembly checkpoint proteins in that it is not required to recruit Mad2 to unattached kinetochores, where Mad2 is believed to form complexes with Cdc20 and BubR1. So how else could USP44 function at the checkpoint? It turns out that, *in vitro*, USP44 can inhibit the ability of UbcH10 to activate checkpoint-inhibited APC/C, leading Elledge and colleagues to propose that USP44 might stabilize Cdc20–Mad2–BubR1 complexes by destroying the ubiquitin chains that APC/C adds to Cdc20 (Fig. 1b). Consistent with this argument, depletion of USP44 prematurely inactivates the spindle-assembly checkpoint in mitotic cells and leads to defects in chromosome segregation.

The model proposing that the stability of Cdc20–Mad2–BubR1 complexes is controlled by a fine balance between ubiquitination, mediated by the APC/C, and de-ubiquitination, catalysed by USP44, makes a number of predictions. Testing these will be an essential goal for the future. For example, could a mutant of Cdc20 be created that couldn't be ubiquitinated but would otherwise be functional? If so, such a mutant would be predicted to assemble into unusually stable checkpoint complexes from which Mad2 and BubR1 could not easily dissociate.

These studies[1,2] also raise a number of other questions. Is de-ubiquitinating Cdc20 the main role of USP44 in maintaining the spindle-assembly checkpoint, or does it also antagonize APC/C more directly by disassembling ubiquitin chains on its other protein substrates such as securin and cyclin B? How does ubiquitination dissociate Mad2 and BubR1 from Cdc20 — by inducing conformational changes in these proteins, or by recruiting enzymes (such as the p97/Cdc48–ATPase) that would catalyse the dissociation process? Finally, how is the balance between de-ubiquitination and ubiquitination reactions tipped once all chromosomes have become attached to both poles of the mitotic spindle? Elledge *et al.* found that USP44 itself is degraded at the end of cell division. Could this be the primary switch for checkpoint inactivation, or is it merely a consequence of APC/C activation once the checkpoint has been silenced? Answering these questions will keep researchers busy for some time to come. ■

Jan-Michael Peters is at the Research Institute of Molecular Pathology (IMP), Dr. Bohr-Gasse 7, A-1030 Vienna, Austria.
e-mail: peters@imp.univie.ac.at

1. Stegmeier, F. *et al. Nature* **446,** 876–880 (2007).
2. Reddy, S. K., Rape, M., Marganski, W. A. & Kirschner, M. W. *Nature* **446,** 921–925 (2007).
3. Rieder, C. L., Schultz, A., Cole, R. & Sluder, G. *J. Cell Biol.* **127,** 1301–1310 (1994).
4. Nasmyth, K. *Cell* **120,** 739–746 (2005).
5. Li, Y., Gorbea, C., Mahaffey, D., Rechsteiner, M. & Benezra, R. *Proc. Natl Acad. Sci. USA* **94,** 12431–12436 (1997).
6. Hwang, L. H. *et al. Science* **279,** 1041–1044 (1998).
7. Kim, S. H., Lin, D. P., Matsumoto, S., Kitazono, A. & Matsumoto, T. *Science* **279,** 1045–1047 (1998).
8. Peters, J.-M. *Nature Rev. Mol. Cell Biol.* **7,** 644–656 (2006).
9. Sudakin, V., Chan, G. K. & Yen, T. J. *J. Cell Biol.* **154,** 925–936 (2001).
10. Luo, X., Tang, Z., Rizo, J. & Yu, H. *Mol. Cell* **9,** 59–71 (2002).
11. Sironi, L. *et al. EMBO J.* **21,** 2496–2506 (2002).

Correction
In the News & Views article "Cell biology: Lost in mitotic translation" by Anthony Wynshaw-Boris (*Nature* **446,** 274–275; 2007), statements in the text and the caption to Figure 1, and part **b** of Figure 1, imply that eIF4B binds directly to the 5′ cap of mRNA. Rather, eIF4B facilitates the ATP-dependent helicase activity of eIF4A to promote the ribosome recruitment necessary for cap-dependent translation.

OBITUARY

Donald E. Osterbrock (1924–2007)

Astrophysicist and historian of astronomy.

The internal structure and physics of stars such as our Sun, the physical conditions of the gas surrounding many hot stars in our Galaxy, and the workings of so-called active galactic nuclei are diverse areas of modern astrophysical research. Donald Osterbrock, who died on 11 January, made fundamentally important contributions to all of them, tackling their central questions with theoretical understanding and observational skill.

Osterbrock was born in Cincinnati, Ohio, on 13 July 1924, the son of an electrical engineering professor. By high school, he had decided that he wanted to be an astronomer. After army service in the Second World War, he attended the University of Chicago, which was at the time not only a leading centre of physics research, but also possessed of a distinguished faculty of astronomers. The solid grounding in theory and observation that Osterbrock received from such renowned names as Enrico Fermi, Gregor Wentzel, Subrahmanyan Chandrasekhar and Otto Struve was the foundation of his subsequent research career. His doctoral thesis, supervised by Chandrasekhar at Chicago's Yerkes Observatory, was a theoretical study of the dynamics of stars in galaxies as they interact gravitationally with a highly variable medium of stars, gas and dust.

After completing his PhD in 1952, Osterbrock spent a year as a postdoctoral fellow at Princeton University, where the internal structure and evolution of stars was the subject of intensive study. He became interested in low-mass stars similar to the Sun, called red dwarfs, whose radii and luminosities were not correctly predicted by theoretical models of the time. Osterbrock reconciled theory and observation by proposing that it was convective energy transport, associated with bulk turbulent motions of gas — rather than heat radiation as was generally assumed — that transported the energy in the outer portions of these stars. Inspired by this success, Martin Schwarzschild and Fred Hoyle, another visitor to Princeton at this time, used the convection model to compute the first successful models of red-giant stars.

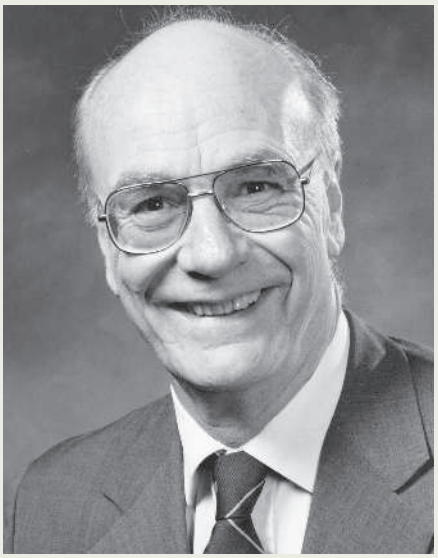
In 1953, Osterbrock was invited to join the nascent astronomy department at the California Institute of Technology (Caltech), which had recently acquired a 200-inch-aperture telescope. Osterbrock's early observational research there included studies of the evolution of stars, both in 'globular clusters' found in the Milky Way and in nearby galaxies. But the future direction of his research was determined around this time by a paper from the theoretical astronomer Michael Seaton. This contained calculations of the interaction probabilities for electrons and oxygen ions, and Osterbrock used the theory to embark on an extended study of O^+ and other ionic emissions from cosmic regions of ionized gas, or nebulae. In a short time, he had established himself as one of the world's leading authorities on gaseous nebulae.

At Caltech, Osterbrock started supervising what was to become a long list of PhD students, many of whom went on to distinguished careers — four became directors of major observatories. In 1958, he moved to help create a first-class department of astronomy at the University of Wisconsin, Madison. Osterbrock and his students were able to use the modern equipment there to continue his studies of gaseous nebulae.

The discovery in the early 1960s of quasars — a particularly luminous class of active galactic nucleus (AGN) radiating considerable amounts of energy from non-stellar sources — intensified interest in the study of external galaxies. The spectra of AGNs are often dominated by the same emission lines as those seen in gaseous nebulae, despite the physical conditions being clearly very different. Osterbrock was thus in an ideal position to unravel the mysteries of the poorly understood AGNs. Such research dominated the rest of his career, and he wrote a classic textbook on the subject, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*, now in its third edition.

In 1973, Osterbrock became director of the Lick Observatory, headquartered on the University of California's Santa Cruz campus. One attraction held by the observatory was a 3-metre-aperture telescope fitted with a highly advanced electronic spectrograph that was ideally suited for Osterbrock's AGN research. During this time, assisted by a string of excellent students, he considerably expanded the classification scheme for different types of AGN, and also steadily improved physical and geometrical models of their structure.

Although Osterbrock carefully reserved about half of his time for research, he took his directorial responsibilities very seriously. During his tenure, the Lick Observatory went through a phase of rapid advances in instrumentation, culminating in the project to build a 10-metre-aperture telescope on a 'dark' site farther away from human light pollution. Two competing designs emerged, one with a primary mirror made of a single, thin meniscus of glass, and the other with



D. HARRIS/UNIV. CALIFORNIA SANTA CRUZ

a primary mirror made up of a mosaic of segments. The meniscus mirror would have been fairly straightforward to produce, whereas the mosaic mirror was more challenging, but weighed less. Each approach had strong supporters, and Osterbrock skilfully led the process that arrived at a choice: a segmented-mirror telescope. The result was the construction of the two Keck telescopes on Mauna Kea, Hawaii, the largest optical/infrared telescopes presently available.

Osterbrock felt uncomfortable with the administrative demands of leading the Lick Observatory when a major new observatory was being built, and he stepped down as director in 1981 to return to full-time teaching and research. In his later years, and especially following his retirement in 1991, he made a second name for himself as a historian of astronomy, writing several highly regarded biographies and histories of the Lick and Yerkes observatories. In 2004, a three-day symposium in honour of Osterbrock's eightieth birthday attracted astronomers and historians of the subject from far afield to Santa Cruz, many of them former students and colleagues. Don Osterbrock was the most sociable of scientists, and the animated discussions at the gathering's coffee breaks and lunches, with him at the centre and frequently punctuated with laughter, will be remembered by the unusually large number of close friends he had gained throughout the world. ■

Joseph S. Miller

Joseph S. Miller, formerly director of University of California Observatories/Lick Observatory, is at the Lick Observatory, University of California, Santa Cruz, California 95064, USA.
e-mail: miller@ucolick.org

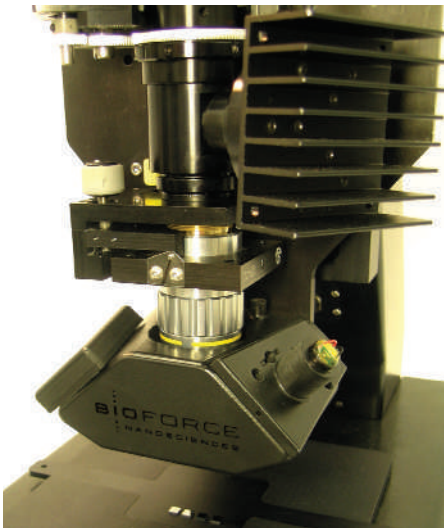
Close-up on cell biology

To define the workings of cellular structures and molecules requires cutting-edge technology not only in biology and biochemistry, but also now in nanotechnology. Hayley M. Birch and Julie Clayton report.

The dawn of the nano-era has placed a whole new array of tools in the hands of cell biologists who are keen to go deeper into the intricacies of how cells work. Forever pushing the boundaries, cell biologists are shifting focus from the micro- towards the nano- and even sub-nano level. To do so means having to find ever more creative ways of using not just biology, but also elements of physics, materials science and engineering. Biophysicists, nanotechnologists, nanofabricators and electrical engineers are working side by side to probe further into the cell.

Nanotopography

Knowing what makes a cell tick is as much about understanding its outside environment as its internal workings. The availability of sophisticated nanopatterning and nanotopography techniques — in which surfaces can be etched and coated with a variety of substances at the nanoscale — is enabling cell biologists to gain a more precise understanding of, and control over, cellular responses compared with cell culture in conventional glass or polystyrene culture plates A range of substrate structures designed specifically for exploring cell behaviours such as spreading, migration and adhesion are under development (see "Cell culture in three dimensions" and "Down to the letter", page 940).



BIOFORCE

The Nano eNabler deposits molecules at nanometre resolution.

The field is so new that most research teams are creating their own nanopatterned surfaces on which the cells are cultured. At Columbia University, New York, Michael Sheetz is developing nanopatterning methods with a team of cell biologists, systems biologists and nanofabricators. "We put down arrays of different spacings of molecules and use those to measure cellular responses," he explains. "We have found that the spacing really matters." Sheetz and colleagues are currently using patterning to examine how the spacing of protein ligands on a substrate surface is important in the binding of dimeric proteins such as talin.

While Sheetz and many others are currently using home-made devices, commercial developers are beginning to sense an opportunity. BioForce Nanoscience in Ames, Iowa, has developed the Nano eNabler, a device that can deposit molecules on surfaces at defined locations at nanometre resolution, and which, among its various options, can be used to apply extracellular matrix proteins onto cell-culture surfaces in predefined patterns. A key feature of the machine is that the drops deposited can be as large as 100 μm or as small as 2 μm

CELL CULTURE IN THREE DIMENSIONS

Taking the idea of nanoscale design for cell culture further, some cell biologists are investigating cellular responses to more complex three-dimensional structures.

A collaboration in this field between cell biologists and electronic engineers at the University of Glasgow's Centre for Cell Engineering is proving fruitful.

Mathis Riehle: "With the E Beam you can be very specific."

The group has the advantage of a state-of-the-art electron-beam nanolithography system (E Beam) manufactured by Leica, which creates nanoscale features on a surface under the control of customized software (see also 'Down to the letter'). "We're quite unique in the United Kingdom in that we have access to such an incredible machine with which you can define structures so deliberately," says Mathis Riehle, who directs research at the centre. "With the E Beam you can be very specific."

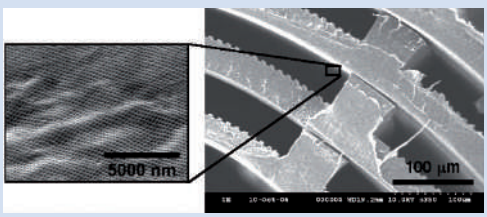
The E Beam is used to define or "write" a pattern at nanometric resolution into an electron-sensing polymer. The pattern is developed and used as a template for etching or depositing material to form the desired nanotopographic features. This technique gives a greater degree of control than 'natural' lithography, which relies on self-assembly of colloidal particles in regular arrays to produce the pattern.

Riehle wants to explore the limits of the machine. "We've made tubes and we want to make more complicated three-dimensional structures — maybe structures that will look like cell 'car parks'. But to do that we would need something like origami, because at the moment the E Beam can only really write on a flat plane. We cannot write on a shaped, undulating surface." The team will have to write its own software to create these complex structures.

As one of a consortium of institutions working on emerging nanopatterning methods, the Glasgow centre is also developing alternative methods of engineering 3D structures with nanoscale features for use as cell-culture substrates.

Kris Seunarine and Osian Meredith have fabricated a 'swiss roll' from e-poly-caprolactone, a biodegradable thermoplastic shaped by hot nano- and micro- embossing. This is the 'sponge' of the swiss roll, with the cells as the 'jam'.

Two levels of microfeatures are sculpted into the polymer (see photos). One aligns fibroblasts and smooth-muscle cells, keeping them from being squashed. The inward-facing surface is patterned at the nanoscale with a regular array of pits (100 nm diameter and 80 nm deep) spaced 300 nm apart, which define locations for endothelial cell adhesion. These cell scaffolds could be useful in developing methods for vascular and urogenital reconstruction. H.M.B.



K. SEUNARINE & O. MEREDITH

Cell support: ridges on the surface of the 'swiss roll' will keep cells aligned, and pits provide anchorage.

VEECO

Two in one: the BioScope II combines atomic force microscopy with fluorescence microscopy.

across — smaller than a single cell. "The cells can interact with many spots at once, which enables the investigator to construct more complex questions, such as how well a cell will respond given a choice of two different proteins," says BioForce's product manager Michael Lynch. The fine control is achieved through a microfluidic cantilever, which dispenses liquid onto a surface that is etched into defined regions.

The Nano eNabler was originally developed for the biosensor market, to look at interactions between proteins, DNA and RNA without the need for labelling. But interest has grown in the past year from researchers wishing to use the device for nanopattern-based cell culture. Around a dozen research groups are currently exploring its potential for this kind of work, including the study of neurons.

Lynch anticipates that in future cell biologists will prefer to use the equipment to design their own nanopatterns rather than buy nanopatterns off-the-shelf. "There are so many different questions being asked that it's hard to narrow down to one pattern, chip or substrate for cell culture. It's not as standard as the microarray field, which has turned into a service."

Two worlds collide

Until recently, cell biologists using confocal microscopy and those using electron microscopy (EM) inhabited two completely separate worlds: confocal microscopists made movies in glorious technicolour, whereas EM practitioners produced black-and-white stills. Both caught cells on camera, but neither got all the results they really wanted.

For confocal film-makers the problem is the lack of resolution, as this has to be sacrificed in order to capture live-action cell behaviour. EM experts, on the other hand, have all the resolution they require, but only a slim chance of snapping a cell in the process of doing something interesting.

But at last these two worlds are slowly converging to give correlative light and electron microscopy through a new and surprisingly simple technology. Ian Lamswood, marketing manager for Leica Microsystems in Vienna, Austria, explains how Leica was approached by Paul Verkade and his colleagues from the Max Planck Institute for Cell Biology in Dresden, Germany. "They wanted to look at the same cell in the EM as they had in the confocal; basically to freeze-frame it and look at the same proteins being synthesized at the same moment in time."

Verkade, an electron microscopist now at the University of Bristol, UK, recalls, "I saw a paper in *Nature Cell Biology Reviews* that said it would be great to have something that could combine fluorescence and electron microscopy. It was a really funny coincidence because we were working on it, but of course, I couldn't say." His invention, currently marketed by Leica as the Rapid Transfer System (RTS), takes a sample from under the confocal microscope to a high-pressure freezer within a few seconds, in preparation for electron microscopy. The RTS has been designed as an attachment to Leica's EM PACT high-pressure freezer. It consists of a rapid loader, which holds the live sample beneath the confocal, then transfers it quickly to the high-pressure freezer, creating a specimen that can be used for EM. Confocal microscopy allows the user to pinpoint a potentially interesting cell event, while the high-pressure freezing preserves the fine ultrastructure at the chosen moment. This provides the researcher with a snapshot of a known cell activity. Using the RTS system, the time for transfer between light microscopy and the fixation of a cell for EM is reduced from around a minute to less than 5 seconds. The state of the cell viewed in the EM is then the same as that last seen by the confocal microscope, but at a much higher resolution. "This machine has the added advantage that it is really good for standard fixation for EM," notes Verkade.

Paul Verkade: solving the problem of combining fluorescence and electron microscopy.

For other investigators, the most appealing new development is the combination of atomic force microscopy (AFM) with fluorescence microscopy. Veeco, of Woodbury, New York, has combined AFM with an inverted optical or confocal microscope in the BioScope II, which enables the high-resolution images of AFM to be complemented by fluorescence microscopy. The BioScope II is an advance over previous Veeco models, with new software for greater flexibility and control by the investigator.

Dennis Discher and his team at the University of Pennsylvania in Philadelphia are using the BioScope II for investigating the differentiation of stem cells on different types of substrates. In particular, they want to assess the rigidity of the cytoskeleton and mechanical changes in its microenvironment at the various stages of differentiation. AFM, like scanning EM, gives highly precise topographical information — what Discher calls "blobology" — such as height, length and shape. But in contrast to EM, AFM can be performed on live cells. Combining this with a fluorescence microscope allows investigators at the same time to identify and track the various structures observed by AFM, for example, in cells expressing proteins tagged with green fluorescent protein.

"You can do mechanical interrogations on cells, and on isolated molecules and complexes, that you can't do with EM — such as make force measurements — and as well you can treat with drugs, and then push and poke to characterize remodelling, and this can all be done before, during and after imaging by fluorescence," says Discher.

JPK Instruments in Berlin, Germany, also markets a new AFM instrument — the NanoWizardII BioAFM — which can be mounted either on a confocal laser-scanning microscope (such as the LSM 510, from Carl Zeiss in Jena, Germany) or on an inverted light microscope (such as the Zeiss Axiovert 200). This allows the user to identify the location of cells of interest across a wide field of view before using AFM to scan a particular region at high resolution and/or perform fluorescence imaging. JPK Instruments also supply a cell-culture vessel (the JPK BioCell) in which cells can be maintained in buffer or cell-culture medium at 37 °C for the duration of the observation. The AFM tip can be coated with ligands or used to prod cells mechanically, and the ensuing biochemical reaction observed by fluorescence.

JPK's DirectOverlay software includes calibration that enables the user to integrate the various types of images and precisely superimpose a fluorescence image onto a 3D projection of a cell. These techniques are particularly

well suited to studying the role of the cytoskeleton in processes such as cell spreading and attachment. Another company improving the performance of atomic-force microscopy by integrating it with confocal microscopy is Asylum Research of Santa Barbara, California, while nAmbition in Dresden, Germany, offers instruments for automated force spectroscopy.

Nanomanipulations

Going beyond observation to direct cell manipulation, Iva Tolic-Nørrelykke and her colleagues at the Max Planck Institute for Cell Biology in Dresden have brought optics giant Olympus in Hamburg and optoelectronics company PicoQuant, based in Berlin, together to market a combination of confocal microscopy and a picosecond (10^{-12} s) pulsed diode laser cutter, to enable manipulation of subcellular structures at the same time as viewing the results. "People have done laser ablation before but mainly with an ultraviolet laser on other microscopes. The new picosecond laser is simple and cheap, and induces less damage than UV because of its longer wavelength and short pulses," Tolic-Nørrelykke explains. She is using the device to examine the effects of laser cutting on the ability of microtubules to maintain the position of cellular organelles.

And taking the idea of combinations a step further, Tolic-Nørrelykke is also exploring, so far in the lab rather than commercially, putting together a two-photon microscope with a femtosecond (10^{-15} s) pulsed laser — to do both fluorescence imaging and laser cutting — and adding another, continuous, laser. This laser will act as optical tweezers, which can trap and displace microscopic objects. (Two-photon microscopes are a type of fluorescence microscope that enables imaging of living tissue down to a depth of 1 millimetre with low bleaching and photodamage.) This combination has enabled her team, for example, to specifically destroy small portions of the cytoskeleton and then use optical tweezers to displace the nucleus and follow how the cytoskeleton works to restore the correct position of the nucleus in the centre of the cell.

The Max Planck team has put together its system from standard commercial products from companies such as Zeiss, Hamamatsu and Thorlabs. "These techniques have existed for 10 to 20 years but we have optimized and combined them in one set-up," says Tolic-Nørrelykke.

Looking ahead, optical tweezers are likely to become an even more sophisticated tool for manipulating either whole cells or subcellular structures. This is being made possible through the use of holograms to split the optical tweezer laser into multiple beams, each one capable of directing the movement of a separate structure. By projecting a sequence of holograms, each positioned slightly differently to the one before, a sequence of light patterns occurs that can be made to move multiple objects around in three dimensions, in a kind of animated dance, according to physicist David Grier at New York University. Grier created the spin-off company Arryx in Chicago, Illinois, based on this concept, the result being BioRyx 200, a device that has a variety of applications, including sorting cells according to size, shape and composition, differentiating between cancerous and non-cancerous cells, and distinguishing bacteria from viruses.

MPI-CBG, DRESDEN

Iva Tolic-Nørrelykke: putting lasers and microscopes together in new combinations to manipulate cellular components.

Further developments in software and photonics means that holographically controlled optical traps will become an important technology for manipulating structures inside cells in the next 10–20 years, predict biologist Daniel Robert and physicist Mervyn Miles at the University of Bristol's new nanoscience centre. Robert and Miles collaborate with Miles Padgett at the University of Glasgow to explore the use of 'haptics' — computer-tracked hand movements — to manipulate objects through the movements of optical traps (see R. Webb, *Nature* **444,** 1017; 2006).

These developments will help cell biologists gain a different perspective. "The mechanical aspects of biology at the micro- and nanoscale have not been explored to their full potential. There are now lots of new tools that can enable us to do that," says Robert.

Nanoprobes

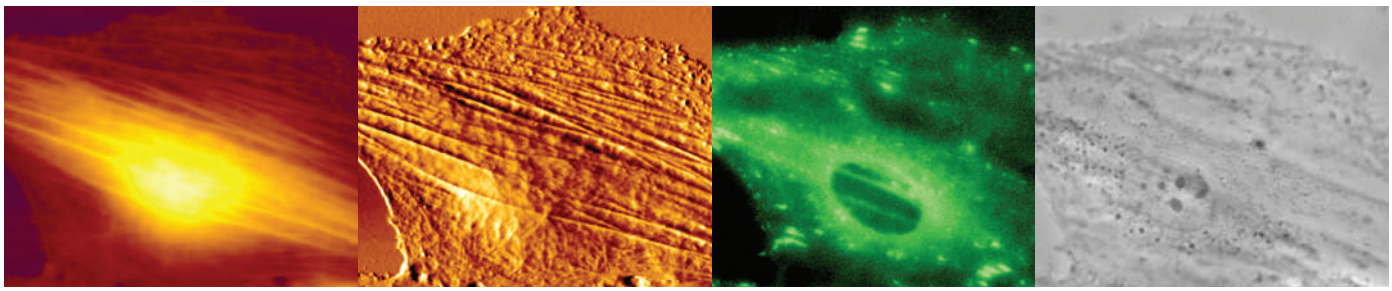
Nanotechnology is also proving useful to biologists wishing to focus on single molecules. Nanoprobes such as quantum dots, semiconducting crystals that can fluoresce in a wide variety of colours, work as cell spies, shadowing the movements of their molecular quarries and allowing the investigator to follow in-cell events on camera with an astonishing level of detail.

Such close camera work requires sophisticated imaging solutions. In confocal microscopy, the majority of photons emitted from a fluorescent light source are rejected, so as few as two out of every 100 falling onto a pixel are lost. Charge-coupled device (CCD) image detectors improve on this low level 'quantum efficiency' by capturing more photons and converting them into electrons, but what is really needed is a way of amplifying the signal above the level of background noise. Colin Coates, market development manager for Andor in Belfast, UK, explains how electron multiplying CCD (EMCCD) improves on standard CCD, "It's like trying to find something hidden in long grass," says Coates, "With EMCCD we amplify to such a high level that those photons aren't lost any more."

Having pioneered the technology, Andor now markets a range of EMCCD cameras for studying single-molecule dynamics. Stefan Diez, at the Max Planck Institute of Cell Biology and Genetics in Dresden, uses an Andor iXon EMCCD in combination with quantum-dot probes 10–20 nm in diameter to watch motor proteins walk on microtubules.

Not content with just observing, Diez is harnessing the power of motor proteins to perform highly specific manipulations, including stretching single DNA molecules as well as networks of DNA. "It's like body surfing," says Diez. "The microtubules are gliding across the surface of the kinesin motors. It's the force of the motors that moves the DNA attached to the microtubules." These kinds of manipulations are highly effective because they employ the cell's own machinery, he explains. "This is actually a good example, not of where nanotechnology helps us to improve biology, but of where biology has been used to improve nanotechnology."

These manipulations could reveal more about DNA mechanics and DNA–enzyme interactions, and also show the potential of



JPK

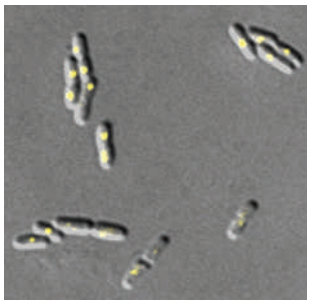
Aspects of a cell: a living fibroblast imaged by atomic force microscopy (first panel, height image; second panel, deflection image) and simultaneously by epifluorescence (third image), and by phase-contrast microscopy (far-right image).

biological motors for use in the molecular manufacturing of nanoelectronic circuits.

Biochemists generally have to break open and extract the contents of cells to study reactions of interest, often after having to synchronize the cells' activity to get a high enough concentration of the molecules of interest in the same state. But now, thanks to a handful of pioneering research groups, it is becoming possible to study individual protein, DNA or RNA molecules in live cells, in real time, and to follow their interactions and kinetics with nanometre-scale spatial precision and millisecond time resolution.

One at a time

Chemical biologist Sunney Xie and his team at Harvard University in Cambridge, Massachusetts, have developed methods for observing the expression of individual protein molecules. After initially testing them in bacterial cells, the group is now extending the techniques to mammalian cells. This approach is useful for detecting proteins expressed at low copy number, at just a few copies per gene, such as transcription factors, which would be undetectable using conventional proteomics methods.



J. ELF & G.-W. LI, HARVARD UNIV.

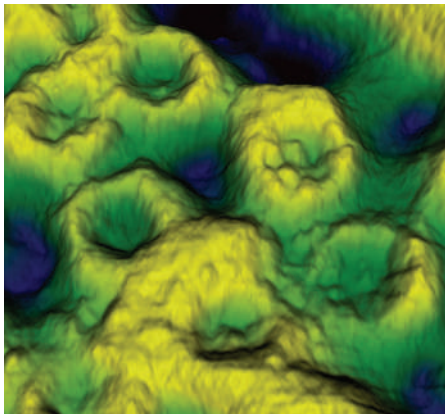
One at a time: single fluorescently labeled proteins binding at one site per *Escherichia coli* chromosome.

The great advantage of the single-molecule method is that individual proteins are detected as they are synthesized in an individual cell. One of Xie's techniques detects the expression of proteins genetically tagged with the fast-maturing yellow fluorescent protein Venus. The appearance of the protein, observed as a series of fluorescent bursts that can be quantified, indicates that transcription and translation are taking place.

Rapid strobing illumination enables the observation of fast binding and unbinding of fluorescently tagged proteins to immobile components in the cell and of interactions between fluorescent proteins in the cytoplasm. "Like photographic images of a bullet going through an apple, we can detect fluorescent proteins in the cytoplasm. We can play with the pulse width to determine the dynamics and where the proteins are," says Xie.

This technology could be extended to a high-throughput system, according to Xie's postdoc Nir Friedman. "You could put cells on a large chip with many chambers and do these kinds of measurements on a large scale."

The fluorescent protein fusion could be done with a DNA library in order to detect the expression of proteins of unknown function. "Although you need to target specifically, the advantage is that you can look at live cells in real time and with single-molecule sensitivity," Friedman adds.



B. WINDOFFER, UNIV. MUNSTER/JPK

A 3D image of a nuclear pore complex.

Nanotechnology seems destined to leave a lasting legacy for cell biology with a host of innovative new technologies. With continuing efforts to combine existing technologies in novel ways, and to create new ones, the possibilities for gaining new insights through nanoscale cell manipulation are increasing rapidly. Nanopatterning and nanotopography are techniques that are, as yet, practised by only a handful of specialists, but the equipment and software are fast becoming available commercially. This trend towards the increasing use of nanotechnology is pushing the very boundaries of cell biology. ■

Hayley M. Birch and Julie Clayton are science writers based in Bristol, UK.

DOWN TO THE LETTER

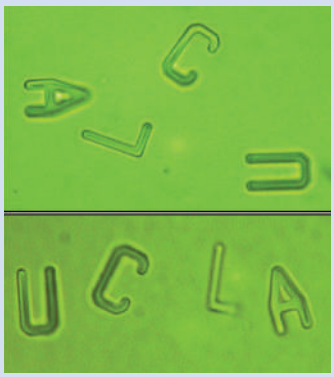
Nanoprobes come in all shapes and sizes. In the latest advance in probe engineering, chemists, physicists and engineers at the University of California, Los Angeles, are pooling their resources to perfect a method for mass-producing novel fluorescent microparticles. The nature of these particles can be so precisely controlled that researchers have been experimenting by creating entire alphabets that can be manipulated with optical tweezers, raising the intriguing possibility of playing nano-scrabble.

These so-called LithoParticles are sculpted by electron-beam lithography, directed by the same computer-aided design (CAD) software used by architects. "E-beam writing is a serial process," says Thomas Mason, who leads the group. "Each letter is written one at a time, so it's not very good for mass production. However, once the mask is made, it can be used over and over again in a special optical-projection printer. We use a mask made by E-beam lithography to expose resist-coated wafers to patterned ultraviolet light. A different projection-printing device — an optical lithography system known as a stepper — is used to mass-produce many particles in parallel." The Ultratech XLS stepper has a lens weighing over 90 kilograms and its own heating and air-conditioning systems to control thermal expansion. The same technology could be used to mass-produce particles with feature sizes as small as 30 nm.

The potential implications for cell biology are huge. Such accuracy of design, coupled with high fidelity on a mass scale, means researchers could soon be supplied with solutions of probes tailored to their specific needs, as neatly demonstrated by Mason's 'alphabet soup'.

Nanoprobes are being increasingly used in the emerging field of bio-microrheology, which examines transport processes within living cells, and in investigating the mechanical properties of cellular components. Nanoparticles introduced by ballistic injection have revealed how the cytoplasm of human umbilical vein endothelial cells undergoes elastic changes in response to growth factors. But the approach could be expanded to investigate the cell's response to all manner of different shapes. "Tracking how differently shaped particles move and rotate inside cells may provide a wealth of information about life cycles and internal cytoplasmic transport in different cell types," says Mason. "You could also use these probes to study how cells respond to various external stimuli. For instance, particles that have many long 'arms' may behave very differently to the compact spheres and quantum dots that are currently available."

UCLA is currently applying to patent their technology and are involved in discussions with commercial partners. Mason is already speculating about building functional nanomachines — including motors, pumps and entire engines — which could be sent to probe even further into the workings of the cell. H.M.B.



J. N. WILKING/T. G. MASON/UCLA

Under the spell: nano-alphabet.

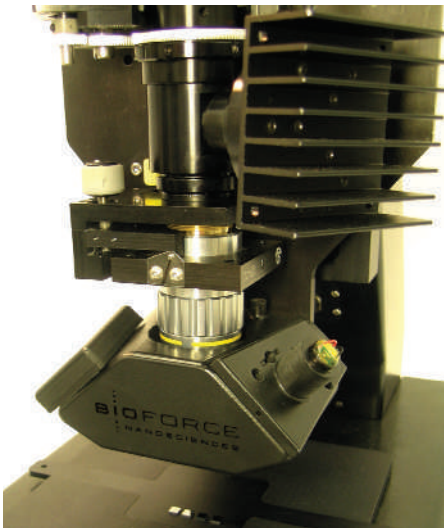
Close-up on cell biology

To define the workings of cellular structures and molecules requires cutting-edge technology not only in biology and biochemistry, but also now in nanotechnology. Hayley M. Birch and Julie Clayton report.

The dawn of the nano-era has placed a whole new array of tools in the hands of cell biologists who are keen to go deeper into the intricacies of how cells work. Forever pushing the boundaries, cell biologists are shifting focus from the micro- towards the nano- and even sub-nano level. To do so means having to find ever more creative ways of using not just biology, but also elements of physics, materials science and engineering. Biophysicists, nanotechnologists, nanofabricators and electrical engineers are working side by side to probe further into the cell.

Nanotopography

Knowing what makes a cell tick is as much about understanding its outside environment as its internal workings. The availability of sophisticated nanopatterning and nanotopography techniques — in which surfaces can be etched and coated with a variety of substances at the nanoscale — is enabling cell biologists to gain a more precise understanding of, and control over, cellular responses compared with cell culture in conventional glass or polystyrene culture plates A range of substrate structures designed specifically for exploring cell behaviours such as spreading, migration and adhesion are under development (see "Cell culture in three dimensions" and "Down to the letter", page 940).



BIOFORCE

The Nano eNabler deposits molecules at nanometre resolution.

The field is so new that most research teams are creating their own nanopatterned surfaces on which the cells are cultured. At Columbia University, New York, Michael Sheetz is developing nanopatterning methods with a team of cell biologists, systems biologists and nanofabricators. "We put down arrays of different spacings of molecules and use those to measure cellular responses," he explains. "We have found that the spacing really matters." Sheetz and colleagues are currently using patterning to examine how the spacing of protein ligands on a substrate surface is important in the binding of dimeric proteins such as talin.

While Sheetz and many others are currently using home-made devices, commercial developers are beginning to sense an opportunity. BioForce Nanoscience in Ames, Iowa, has developed the Nano eNabler, a device that can deposit molecules on surfaces at defined locations at nanometre resolution, and which, among its various options, can be used to apply extracellular matrix proteins onto cell-culture surfaces in predefined patterns. A key feature of the machine is that the drops deposited can be as large as 100 μm or as small as 2 μm

CELL CULTURE IN THREE DIMENSIONS

Taking the idea of nanoscale design for cell culture further, some cell biologists are investigating cellular responses to more complex three-dimensional structures.

A collaboration in this field between cell biologists and electronic engineers at the University of Glasgow's Centre for Cell Engineering is proving fruitful.

Mathis Riehle: "With the E Beam you can be very specific."

The group has the advantage of a state-of-the-art electron-beam nanolithography system (E Beam) manufactured by Leica, which creates nanoscale features on a surface under the control of customized software (see also 'Down to the letter'). "We're quite unique in the United Kingdom in that we have access to such an incredible machine with which you can define structures so deliberately," says Mathis Riehle, who directs research at the centre. "With the E Beam you can be very specific."

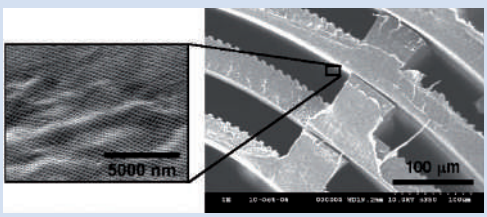
The E Beam is used to define or "write" a pattern at nanometric resolution into an electron-sensing polymer. The pattern is developed and used as a template for etching or depositing material to form the desired nanotopographic features. This technique gives a greater degree of control than 'natural' lithography, which relies on self-assembly of colloidal particles in regular arrays to produce the pattern.

Riehle wants to explore the limits of the machine. "We've made tubes and we want to make more complicated three-dimensional structures — maybe structures that will look like cell 'car parks'. But to do that we would need something like origami, because at the moment the E Beam can only really write on a flat plane. We cannot write on a shaped, undulating surface." The team will have to write its own software to create these complex structures.

As one of a consortium of institutions working on emerging nanopatterning methods, the Glasgow centre is also developing alternative methods of engineering 3D structures with nanoscale features for use as cell-culture substrates.

Kris Seunarine and Osian Meredith have fabricated a 'swiss roll' from e-poly-caprolactone, a biodegradable thermoplastic shaped by hot nano- and micro- embossing. This is the 'sponge' of the swiss roll, with the cells as the 'jam'.

Two levels of microfeatures are sculpted into the polymer (see photos). One aligns fibroblasts and smooth-muscle cells, keeping them from being squashed. The inward-facing surface is patterned at the nanoscale with a regular array of pits (100 nm diameter and 80 nm deep) spaced 300 nm apart, which define locations for endothelial cell adhesion. These cell scaffolds could be useful in developing methods for vascular and urogenital reconstruction. H.M.B.



K. SEUNARINE & O. MEREDITH

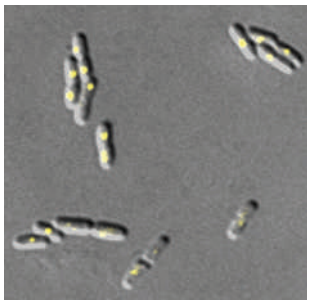
Cell support: ridges on the surface of the 'swiss roll' will keep cells aligned, and pits provide anchorage.

biological motors for use in the molecular manufacturing of nanoelectronic circuits.

Biochemists generally have to break open and extract the contents of cells to study reactions of interest, often after having to synchronize the cells' activity to get a high enough concentration of the molecules of interest in the same state. But now, thanks to a handful of pioneering research groups, it is becoming possible to study individual protein, DNA or RNA molecules in live cells, in real time, and to follow their interactions and kinetics with nanometre-scale spatial precision and millisecond time resolution.

One at a time

Chemical biologist Sunney Xie and his team at Harvard University in Cambridge, Massachusetts, have developed methods for observing the expression of individual protein molecules. After initially testing them in bacterial cells, the group is now extending the techniques to mammalian cells. This approach is useful for detecting proteins expressed at low copy number, at just a few copies per gene, such as transcription factors, which would be undetectable using conventional proteomics methods.



J. ELF & G.-W. LI, HARVARD UNIV.

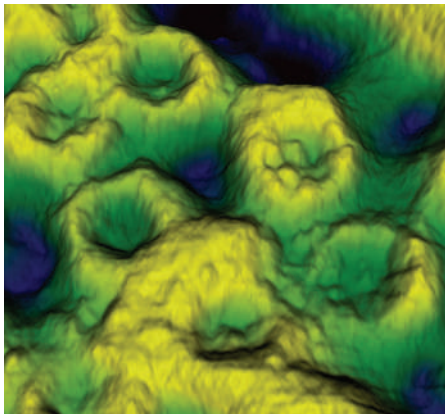
One at a time: single fluorescently labeled proteins binding at one site per *Escherichia coli* chromosome.

The great advantage of the single-molecule method is that individual proteins are detected as they are synthesized in an individual cell. One of Xie's techniques detects the expression of proteins genetically tagged with the fast-maturing yellow fluorescent protein Venus. The appearance of the protein, observed as a series of fluorescent bursts that can be quantified, indicates that transcription and translation are taking place.

Rapid strobing illumination enables the observation of fast binding and unbinding of fluorescently tagged proteins to immobile components in the cell and of interactions between fluorescent proteins in the cytoplasm. "Like photographic images of a bullet going through an apple, we can detect fluorescent proteins in the cytoplasm. We can play with the pulse width to determine the dynamics and where the proteins are," says Xie.

This technology could be extended to a high-throughput system, according to Xie's postdoc Nir Friedman. "You could put cells on a large chip with many chambers and do these kinds of measurements on a large scale."

The fluorescent protein fusion could be done with a DNA library in order to detect the expression of proteins of unknown function. "Although you need to target specifically, the advantage is that you can look at live cells in real time and with single-molecule sensitivity," Friedman adds.



B. WINDOFFER, UNIV. MUNSTER/JPK

A 3D image of a nuclear pore complex.

Nanotechnology seems destined to leave a lasting legacy for cell biology with a host of innovative new technologies. With continuing efforts to combine existing technologies in novel ways, and to create new ones, the possibilities for gaining new insights through nanoscale cell manipulation are increasing rapidly. Nanopatterning and nanotopography are techniques that are, as yet, practised by only a handful of specialists, but the equipment and software are fast becoming available commercially. This trend towards the increasing use of nanotechnology is pushing the very boundaries of cell biology. ■

Hayley M. Birch and Julie Clayton are science writers based in Bristol, UK.

DOWN TO THE LETTER

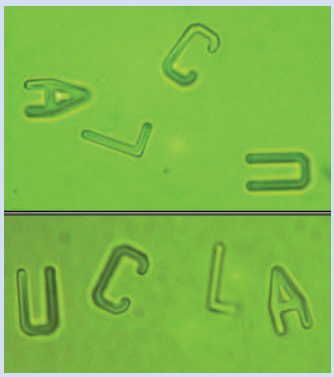
Nanoprobes come in all shapes and sizes. In the latest advance in probe engineering, chemists, physicists and engineers at the University of California, Los Angeles, are pooling their resources to perfect a method for mass-producing novel fluorescent microparticles. The nature of these particles can be so precisely controlled that researchers have been experimenting by creating entire alphabets that can be manipulated with optical tweezers, raising the intriguing possibility of playing nano-scrabble.

These so-called LithoParticles are sculpted by electron-beam lithography, directed by the same computer-aided design (CAD) software used by architects. "E-beam writing is a serial process," says Thomas Mason, who leads the group. "Each letter is written one at a time, so it's not very good for mass production. However, once the mask is made, it can be used over and over again in a special optical-projection printer. We use a mask made by E-beam lithography to expose resist-coated wafers to patterned ultraviolet light. A different projection-printing device — an optical lithography system known as a stepper — is used to mass-produce many particles in parallel." The Ultratech XLS stepper has a lens weighing over 90 kilograms and its own heating and air-conditioning systems to control thermal expansion. The same technology could be used to mass-produce particles with feature sizes as small as 30 nm.

The potential implications for cell biology are huge. Such accuracy of design, coupled with high fidelity on a mass scale, means researchers could soon be supplied with solutions of probes tailored to their specific needs, as neatly demonstrated by Mason's 'alphabet soup'.

Nanoprobes are being increasingly used in the emerging field of bio-microrheology, which examines transport processes within living cells, and in investigating the mechanical properties of cellular components. Nanoparticles introduced by ballistic injection have revealed how the cytoplasm of human umbilical vein endothelial cells undergoes elastic changes in response to growth factors. But the approach could be expanded to investigate the cell's response to all manner of different shapes. "Tracking how differently shaped particles move and rotate inside cells may provide a wealth of information about life cycles and internal cytoplasmic transport in different cell types," says Mason. "You could also use these probes to study how cells respond to various external stimuli. For instance, particles that have many long 'arms' may behave very differently to the compact spheres and quantum dots that are currently available."

UCLA is currently applying to patent their technology and are involved in discussions with commercial partners. Mason is already speculating about building functional nanomachines — including motors, pumps and entire engines — which could be sent to probe even further into the workings of the cell. H.M.B.



J. N. WILKING/T. G. MASON/UCLA

Under the spell: nano-alphabet.

Imaging

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Andor Technology	Scientific imaging and spectroscopy	Belfast, UK	www.andor-tech.com
Apogee Instruments	High-performance cooled CCD imaging systems	Roseville, California	www.ccd.com
Applied Cytometry Systems	Software and integrated cytometry systems	Sheffield, UK	www.appliedcytometry.com
Applied Precision	cellWoRx for imaging living cells	Issaquah, Washington	www.appliedprecision.com
Applied Scientific	Automated systems for microscopy and fluorescence screening	Eugene, Oregon	www.asiimaging.com
Asylum Research	Atomic-force microscopes (AFM), dual AFM and confocal microscope, and accessories	Santa Barbara, California	www.asylumresearch.com
BioForce Nanosciences	Nano eNabler desktop molecular printing system, products for atomic-force microscopy	Ames, Iowa	www.bioforcenano.com
Biomedical Photometrics	MACROscope wide-field confocal laser-scanning microscopes for fluorescence-based assays	Waterloo, Ontario	www.genefocus.com
Bioptonics	Optical projection tomography scanning services	Edinburgh, UK	www.bioptonics.com
Bitplane	Imaris image analysis software for microscopy	Zurich, Switzerland	www.bitplane.com
Cambio	STAR FISH chromosome paints, specialized biochemicals	Cambridge, UK	www.cambio.co.uk
Carl Zeiss	Stereomicrosopes, upright microscopes, inverted microscopes, laser-scanning microscopes, fluorescence correlation spectroscopy, confocal imaging	Gottingen, Germany	www.zeiss.com
Cell Robotics International	LaserTweezers and LaserScissors microscope workstations	Albuquerque, New Mexico	www.cellrobotics.com
Chroma Technology	Optical filters and fluorescence filter sets	Brattleboro, Vermont	www.chroma.com
Clemex	Image analysis software	Longueuil, Quebec	www.celemx.com
Covalys Biosciences	SNAP-tags for labelling cellular components	Witterswil, Switzerland	www.covalys.com
Cyntellect	LEAP automated system for high-throughput cell imaging and laser-based cell manipulation	San Diego, California	www.cyntellect.com
Delong Instruments	LVEM 5 low-voltage electron microscope	Brno, Czech Republic	www.dicomps.com
Evident Technologies	Quantum dot fluorescence labels for molecular interaction assays	Troy, New York	www.evidenttech.com
FEI Company	Electron microscopes and ion-beam technology	Hillsboro, Oregon	www.feicompany.com
Fujifilm	Imaging systems for protein electrophoresis and array analysis	Stamford, Connecticut	www.fujimed.com
Guava Technologies	Benchtop cytometers	Hayward, California	www.guavatechnologies.com
Hamamatsu	Imaging systems	Hamamatsu City, Japan	www.hamamatsu.com
Improvision	3D imaging software	Coventry, UK	www.improvision.com
Intracellular Imaging	InCyt imaging system for imaging calcium and other ions	Cincinnati, Ohio	www.intracellular.com
JPK Instruments	Scanning probe microscopes, NanoWizardII BioAFM dual AFM and confocal microscope	Berlin, Germany	www.jpk.com
Kodak	Imaging systems and image-analysis software	Rochester, New York	www.kodak.com
LaVision Biotec	TriMScope multifocal multiphoton laser-scanning microscope	Bielefeld, Germany	www.lavisionbiotec.com
Leica Microsystems	DN digital microscope range, microscopes and microscope systems	Wetzlar, Germany	www.leica-microsystems.com
LightLab	Optical coherence tomography development	Westford, Massachusetts	www.lightlabimaging.com
Lightools	Whole mouse imaging, fluorescent imaging gel documentation systems, thermal paper	Encinitas, California	www.lightools.com
Mauna Kea Technologies	Miniaturized fibre-optic fluorescence confocal microscopy system for *in vivo* intra-tissue imaging	Paris, France	www.maunakeatech.com
MBL International Corporation	Range of CoralHue proteins	Woburn, Massachusetts	www.mblintl.com
Micro Video Instruments	Microscopes and accessories	Avon, Massachusetts	www.mvi-inc.com
MMI Molecular Machines & Industries	Equipment for micromanipulation and laser microdissection	Heidelberg, Germany	www.molecular-machines.com
Molecular Devices	Liquid-handling and microplate-processing equipment, imaging instruments and assay systems	Sunnyvale, California	www.moleculardevices.com
nAmbition	Instruments for atomic-force spectroscopy	Dresden, Germany	www.nambition.com
Nanoptek	Digital photon tunnelling microscopes	Concord, Massachusetts	www.nanoptek.com
Nikon Instruments	Microscopes, COOLSCOPE digital microscope with screen	Kanagawa, Japan	www.nikon-instruments.jp/eng/
Olympus	Digital microscopes, upright and inverted microscopes, stereo microscopes, confocal laser-scanning microscopes	Hamburg, Germany	www.olympus.de/microscopy
Omega Optical	Filter sets, microscopes	Brattleboro, Vermont	www.omegafilters.com
Optronics	Digital microscope cameras, image-analysis software	Goleta, California	www.optronics.com
PerkinElmer Life Sciences	UltraView confocal laser scanning microscope system for live cell imaging	Boston, Massachusetts	las.perkinelmer.com
Photometrics	CCD cameras	Tucson, Arizona	www.photomet.com
PicoQuant	Optoelectronics, photon-counting instrumentation, pulsed lasers and light sources	Berlin, Germany	www.picoquant.com
Princeton Instruments	Equipment for imaging and spectroscopy	Trenton, New Jersey	www.princetoninstruments.com
QImaging	High-performance digital cameras and RGB filters for microscopy	Burnaby, British Columbia	www.qimaging.com

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
QuantumDot Corporation	Quantum dot nanocrystalline fluorescent tags for antibody imaging	Hayward, California	www.qdots.com
Varian	Spectrophotometers, infra-red microscopes and imaging systems	Palo Alto, California	www.varianinc.com
Veeco	Dual AFM and confocal microscope (BioScopeII), probes for scanning force microscopes	Newbury, New York	www.veeco.com
Cell culture and cell monitoring			
Applied BioPhysics	Automated instruments for cell monitoring and electric cell-substrate impedance sensing	Troy, New York	www.biophysics.com
BD Biosciences	FACS range of flow cytometers, equipment for cell and molecular biology	Franklin Lakes, New Jersey	www.bd.com
Biomol	Services for chemical synthesis, molecular biology, antibody production, cell culture	Hamburg, Germany	www.biomol.de
Cytogration	Automated high-throughput cell culture systems for *in vitro* screening	Rockville, Maryland	www.cytogration.com
EUGENEX Biotechnologies	Development of test cell lines	Taegerwilen, Switzerland	www.eugenex.com
Harvard Apparatus	Instruments and equipment for electrophysiology and cell biology	Holliston, Massachusetts	www.harvardapparatus.com
Integra Biosciences	Equipment and consumables for sterilization, liquid handling, cell culture, sample storage	Baar, Switzerland	www.integra-biosciences.com
Molecular Biologicals International	ECOS competent bacterial cells	Irvine, California	www.growcells.com
Multi Channel Systems	Automated equipment for *Xenopus* injection and ion-channel screening, systems for studying cultured nerve cells and neuroregeneration	Reutlingen, Germany	www.multichannelsystems.com
OptiCell	Automated cell-culture devices	Westerville, Ohio	www.opticell.com
Tecan	Freedom EVO automated liquid-handling workstations, Cellerity cell-culture system	Männedorf, Switzerland	www.tecan.com
General			
Agilent	Instrumentation, kits and consumables for the life sciences	Palo Alto, California	www.agilent.com
Alexis Biochemicals	Suppliers of reagents for cell biology, signal transduction, and molecular biology	Lausanne, Switzerland	www.alexis-corp.com
Applied Biosystems	Automated equipment for molecular and cell biology research	Foster City, California	www.appliedbiosystems.com
Beckman Coulter	Automated tools for molecular biology, biochemistry, genomics and proteomics	Fullerton, California	www.beckmancoulter.com
Bio-Rad	Products, instruments and software for life-sciences research	Hercules, California	www.bio-rad.com
Cambrex	Gels and other products for molecular and cell biology research	East Rutherford, New Jersey	www.cambrex.com
CyBio	Pipetting, liquid handling, incubation, and imaging systems for automated screening	Jena, Germany	www.cybio-ag.com
Merck Chemicals	Kits and consumables for molecular and cell biology research	Darmstadt, Germany	splash.emdbiosciences.com
Eppendorf	Laboratory instrumentation and consumables for molecular and cell biology research	Hamburg, Germany	www.eppendorf.com
GE Healthcare	LEADseeker and IN Cell cellular imaging systems for drug screening	Little Chalfont, UK	www.gehealthcare.com
GenScript	Kits and services for cell biology and molecular biology	Piscataway, New Jersey	www.genscript.com
Gilson	Liquid-handling and pipetting instruments	Middleton, Wisconsin	www.gilson.com
Invitrogen	Equipment and reagents for cell and molecular biology research	Carlsbad, California	www.invitrogen.com
Millipore	Automated equipment for molecular biology, biochemistry, genomics and proteomics	Bedford, Masachusetts	www.millipore.com
MP Biomedicals	Reagents and biochemicals for life-sciences research	Solon, Ohio	www.mpbio.com
Nalge Nunc International	Labware	Rochester, New York	www.nalgenunc.com
Pierce Biotechnology	Components and consumables for cell and molecular biology	Rockford, Illinois	www.piercenet.com
Pro Scientific	Laboratory equipment, centrifuges	Oxford, Connecticut	www.proscientific.com
Qiagen	Laboratory workstations and equipment for molecular and cell biology	Venlo, The Netherlands	www.qiagen.com
R&D Systems	Reagents and assays for molecular and cell biology.	Minneapolis, Minnesota,	www.RnDSystems.com
Roche Applied Science	Reagents and kits for molecular biology, functional genomics and proteomics research	Indianapolis, Indiana	www.roche-applied-science.com
Sigma-Aldrich	Reagents and biochemicals for life-sciences research	St Louis, Missouri	sigmaaldrich.com
Stratagene	Tools and reagents for molecular biology, genomics, proteomics and drug discovery	La Jolla, California	www.stratagene.com
TaKaRa	Bio Reagents, kits and custom services for life-sciences research	Shiga, Japan	www.takara.bio.com
Tocris Cookson	Chemicals for life-science research; custom synthesis and contract research services	Avonmouth, UK	www.tocris.com
Wako Chemicals	Speciality chemicals, bioproducts and clinical diagnostic reagents	Richmond, Virginia	www.wakousa.com

LETTERS

Damage to the prefrontal cortex increases utilitarian moral judgements

Michael Koenigs[1]†*, Liane Young[2]*, Ralph Adolphs[1,3], Daniel Tranel[1], Fiery Cushman[2], Marc Hauser[2] & Antonio Damasio[1,4]

The psychological and neurobiological processes underlying moral judgement have been the focus of many recent empirical studies[1–11]. Of central interest is whether emotions play a causal role in moral judgement, and, in parallel, how emotion-related areas of the brain contribute to moral judgement. Here we show that six patients with focal bilateral damage to the ventromedial prefrontal cortex (VMPC), a brain region necessary for the normal generation of emotions and, in particular, social emotions[12–14], produce an abnormally 'utilitarian' pattern of judgements on moral dilemmas that pit compelling considerations of aggregate welfare against highly emotionally aversive behaviours (for example, having to sacrifice one person's life to save a number of other lives)[7,8]. In contrast, the VMPC patients' judgements were normal in other classes of moral dilemmas. These findings indicate that, for a selective set of moral dilemmas, the VMPC is critical for normal judgements of right and wrong. The findings support a necessary role for emotion in the generation of those judgements.

The basis of our moral judgements has been a long-standing focus of philosophical inquiry and, more recently, active empirical investigation. In a departure from traditional rationalist approaches to moral cognition that emphasize the role of conscious reasoning from explicit principles[15], modern accounts have proposed that emotional processes, conscious or unconscious, may also play an important role[16,17]. Emotion-based accounts draw support from multiple lines of empirical work: studies of clinical populations reveal an association between impaired emotional processing and disturbances in moral behaviour[1–4]; neuroimaging studies consistently show that tasks involving moral judgement activate brain areas known to process emotions[5–9]; and behavioural studies demonstrate that manipulation of affective state can alter moral judgements[10,11]. However, neuroimaging studies do not settle whether putatively 'emotional' activations are a cause or consequence of moral judgement; behavioural studies in healthy individuals do not address the neural basis of moral judgement; and no clinical studies have specifically examined the moral judgements (as opposed to moral reasoning or moral behaviour) of patients with focal brain lesions. In brief, none of the existing studies establishes that brain areas integral to emotional processes are necessary for the generation of normal moral judgements. As a result, there remains a critical gap in the evidence relating moral judgement, emotion and the brain.

Investigating moral judgements in individuals with focal damage to the ventromedial prefrontal cortex (VMPC) provides a key test. The VMPC projects to basal forebrain and brainstem regions that execute bodily components of emotional responses[18], and neurons within the VMPC encode the emotional value of sensory stimuli[19]. Patients with VMPC lesions exhibit generally diminished emotional responsivity and markedly reduced social emotions (for example, compassion, shame and guilt) that are closely associated with moral values[1,2,12–14,16], and also exhibit poorly regulated anger and frustration tolerance in certain circumstances[20,21]. Despite these patent defects both in emotional response and emotion regulation, the capacities for general intelligence, logical reasoning, and declarative knowledge of social and moral norms are preserved[20–23]. We selected a sample of six patients with adult-onset, focal bilateral VMPC lesions (Fig. 1) as well as both neurologically normal (NC) and brain-damaged comparison (BDC) subjects. Importantly, each of the VMPC patients had striking defects in social emotion but generally intact intellect and normal baseline mood (Tables 1 and 2, see also Supplementary Table 1). In particular, all six VMPC patients had impaired autonomic activity in response to emotionally charged pictures (Table 2), as well as severely diminished empathy, embarrassment and guilt (Table 2). All comparison subjects (NC and BDC) had intact emotional processing.

Subjects evaluated moral dilemmas designed to pit two competing considerations against one another. A paradigmatic dilemma of this type presents subjects with the choice of whether or not to sacrifice one person's life to save the lives of others. One consideration is a utilitarian calculation of how to maximize aggregate welfare, whereas the other is a strong emotional aversion to the proposed action. One model holds that endorsement of the proposed action (the utilitarian response) requires the subject to overcome an emotional response against inflicting direct harm to another person (a 'personal' harm[7,8]). If emotional responses mediated by VMPC are indeed a critical influence on moral judgement, individuals with VMPC lesions should exhibit an abnormally high rate of utilitarian judgements on the emotionally salient, or 'personal', moral scenarios (for example, pushing one person off a bridge to stop a runaway boxcar from hitting five people), but a normal pattern of judgements on the less emotional, or 'impersonal', moral scenarios (for example, turning a runaway boxcar away from five people but towards one person). If, alternatively, emotion does not play a causal role in the generation of moral judgements but instead follows from the judgements[24,25], then individuals with emotion defects due to VMPC lesions should show a normal pattern of judgements on all scenarios.

To test for between-group differences in the probability of utilitarian responses given for each scenario type (non-moral, impersonal moral, personal moral), we used a logistic regression fitted with the generalized estimating equations method (Fig. 2). There were no significant differences between groups on the non-moral or impersonal moral scenarios (all P values >0.29, corrected for multiple

[1]Department of Neurology, University of Iowa Hospitals and Clinics, Iowa City, Iowa 52242, USA. [2]Department of Psychology, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Division of Humanities and Social Sciences and Division of Biology, California Institute of Technology, Pasadena, California 91125, USA. [4]Brain and Creativity Institute and Dornsife Center for Cognitive Neuroimaging, University of Southern California, Los Angeles, California 90089, USA. †Present address: National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892-1440, USA.
*These authors contributed equally to this work.

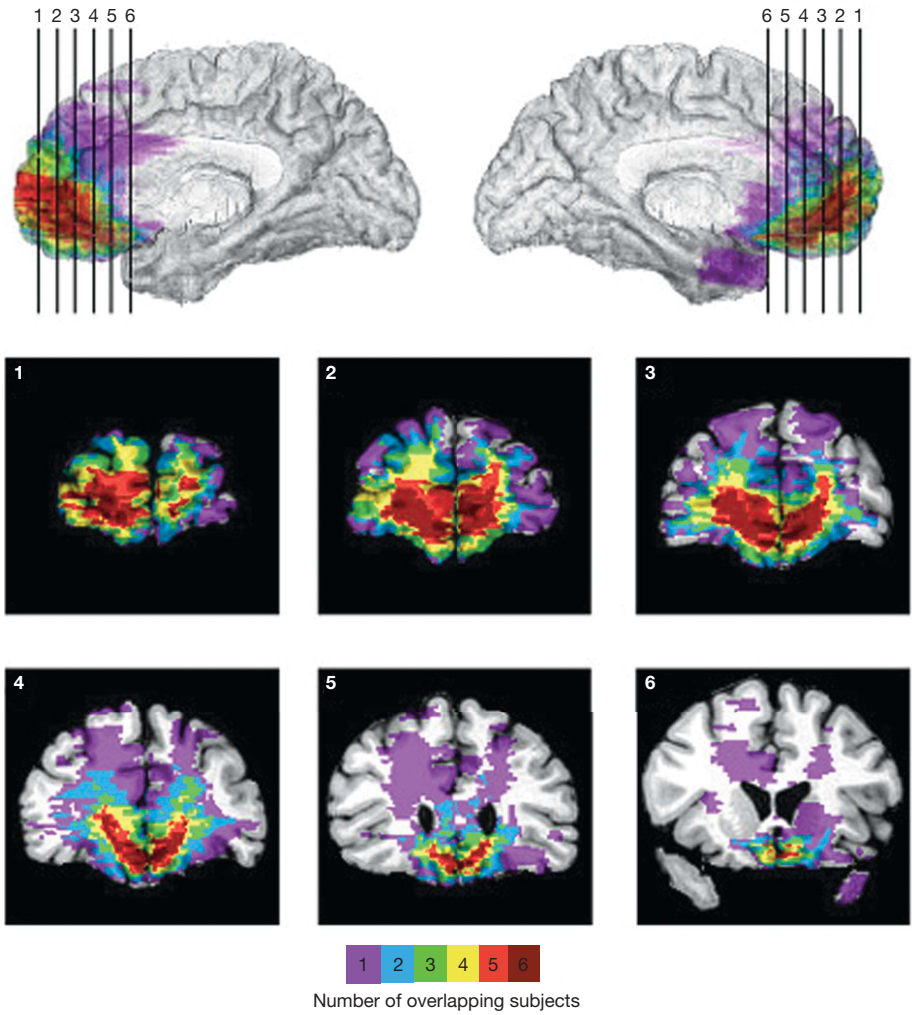


Figure 1 | Lesion overlap of VMPC patients. Lesions of the six VMPC patients displayed in mesial views and coronal slices. The colour bar indicates the number of overlapping lesions at each voxel.

comparisons). In contrast, for personal moral scenarios, the VMPC group was more likely to endorse the proposed action than either the NC group (odds ratio = 2.81; $P = 0.04$, corrected) or BDC group (odds ratio = 3.30; $P = 0.006$, corrected). There was no difference between the NC and BDC groups (odds ratio = 0.85; $P = 0.68$, uncorrected). These data indicate that the VMPC group's responses differed only for personal moral scenarios, suggesting that VMPC-mediated processes affect only those moral judgements involving emotionally salient actions.

In a more fine-grained analysis, we examined response patterns within the personal moral scenarios. For seven out of the 21 personal moral scenarios, both comparison groups were at 100% agreement in their judgements. An additional eighth scenario elicited 100% agreement from the BDC group, and near-perfect agreement from the NC group (with only one participant deviating from the shared response). These eight scenarios were therefore classified as 'low-conflict' (for example, abandoning one's baby to avoid the burden of caring for it). The remaining 13 scenarios (none of which elicited 100% agreement from either comparison group) were classified as 'high-conflict' (for example, smothering one's baby to save a number

Table 1 | VMPC patient neuropsychological data

Subject	WAIS-III			WMS-III		TT	WCST	Stroop	BDI
	VIQ	PIQ	FSIQ	GMI	WMI				
1	142	134	143	109	124	44	6	70	0
2	89	97	91	59	102	44	6	49	3
3	111	96	104	74	105	44	6	67	10
4	108	102	106	109	124	44	6	57	1
5	110	107	109	105	102	44	6	54	8
6	89	80	84	96	88	44	0	77	7

WAIS-III, Wechsler Adult Intelligence Scale-III scores (VIQ, verbal IQ; PIQ, performance IQ; FSIQ, full-scale IQ). WMS-III, Wechsler Memory Scale-III scores (GMI, general memory index; WMI, working memory index). TT, Token Test (from the Multilingual Aphasia Examination), a measure of basic verbal comprehension. WCST, Wisconsin Card Sort Test categories, a measure of executive function. Stroop, *T*-score on the Interference trial of the Stroop Colour-Word Test, a measure of response inhibition. BDI, Beck Depression Inventory, a measure of baseline mood. All patients were within normal ranges except for subjects 2 and 3 on GMI and subject 6 on WCST and Stroop.

Table 2 | VMPC patient social emotion data

Subject	SCRs	Empathy	Embarrassment	Guilt
1	Impaired	3	3	3
2	Impaired	3	3	3
3	Impaired	3	3	3
4	Impaired	2	2	1
5	Impaired	3	3	3
6	Impaired	3	3	3

SCRs, skin conductance responses to emotionally charged socially significant stimuli (for example, pictures of social disasters, mutilations, nudes), using methods previously described[12]. The same SCR experiment was performed in ten of twelve BDC patients, and all ten demonstrated normal SCRs to emotionally charged pictures. A clinical neuropsychologist blind to the hypotheses of the current study rated each VMPC patient's demonstrated capacity for empathy, embarrassment and guilt in his or her personal life. The rating used a four-point scale denoting severity of impairment, where 0 = normal, 1 = mild, 2 = moderate and 3 = severe. Ratings were based on data derived from spouse or family member reports in the Iowa Rating Scales of Personality Change[29] and from data from clinical interviews. Both of these sources provide direct observations about the patient's basic and social emotions, and include questions about whether the patient experiences and manifests emotions such as sadness, anxiety, empathy, embarrassment and guilt.

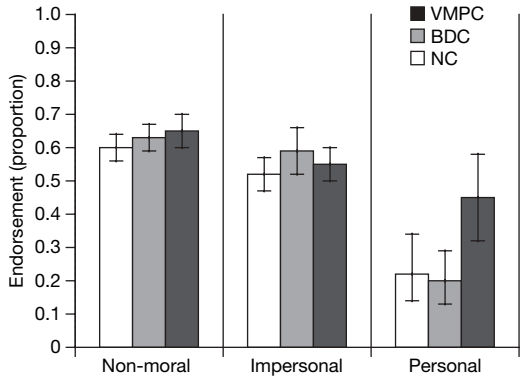


Figure 2 | Moral judgements for each scenario type. Proportions of 'yes' judgements are shown for each subject group. Error bars indicate 95% confidence intervals. We used three classes of stimuli: non-moral scenarios ($n = 18$), impersonal moral scenarios ($n = 11$), and personal moral scenarios ($n = 21$). On personal moral scenarios, the frequency of endorsing 'yes' responses was significantly greater in the VMPC group than in either comparison group (P values < 0.05, corrected).

of people). Reaction-time data support this distinction: response latencies in the NC group on high-conflict scenarios were significantly longer than on low-conflict scenarios (t-test with 19 degrees of freedom, $t(19) = -3.63$; $P = 0.002$).

Like the patients in the comparison groups, the VMPC patients uniformly rejected the proposed action in every one of the low-conflict scenarios (Fig. 3). In contrast, significant differences emerged for the high-conflict scenarios: the VMPC group was more likely to endorse the proposed action than either the NC (odds ratio = 4.70; $P = 0.05$, corrected) or BDC group (odds ratio = 5.38; $P = 0.02$, corrected), with no difference between the NC and BDC participants (odds ratio = 0.87; $P = 0.77$, uncorrected). Every high-conflict personal scenario elicited the same pattern: a greater proportion of the VMPC group endorsed the action than either comparison group.

To recapitulate, VMPC patients' judgements differed from comparison subjects' only for the high-conflict personal moral dilemmas, all of which featured competing considerations of aggregate welfare on the one hand, and, on the other hand, harm to others that would normally evoke a strong social emotion. Low-conflict personal moral scenarios lacked this degree of competition. This difference probably

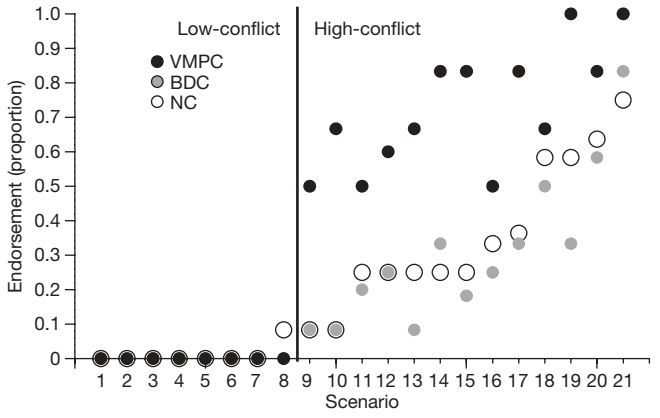


Figure 3 | Moral judgements on individual personal moral scenarios. Proportions of 'yes' judgements given by each subject group for each of the 21 personal moral scenarios. Individual scenarios (numbered 1–21 on the x axis) are ordered by increasing proportion of 'yes' responses given by the normal comparison group. Responses did not differ between subject groups for the low-conflict scenarios (left of the vertical line). The VMPC group made a greater proportion of 'yes' judgements than either comparison group for every one of the high-conflict scenarios (right of the vertical line).

accounts for the greater consensus and faster reaction times on low-conflict personal dilemmas in the comparison groups, and it can also account for the VMPC patients' pattern of judgements. Evidence suggests that knowledge of explicit social and moral norms is intact in individuals with VMPC damage[21,22]. In the absence of an emotional reaction to harm of others in personal moral dilemmas, VMPC patients may rely on explicit norms endorsing the maximization of aggregate welfare and prohibiting the harming of others. This strategy would lead VMPC patients to a normal pattern of judgements on low-conflict personal dilemmas but an abnormal pattern of judgements on high-conflict personal dilemmas, precisely as was observed. The specificity of this result argues against a general deficit in the capacity for moral judgement following VMPC damage. Rather, VMPC seems to be critical only for moral dilemmas in which social emotions play a pivotal role in resolving moral conflict[4,8,16,17].

It is important to note that the effects of VMPC damage on emotion processing depend on context. In this study, the VMPC patients' abnormally high rate of utilitarian judgements is attributed to diminished social emotion, whereas in a recent study of the Ultimatum Game, theVMPC patients' abnormally high rate of rejection of unfair monetary offers was attributed to poorly controlled frustration, manifested as exaggerated anger[20]. These seemingly contradictory findings highlight two distinct aspects of emotion impairment that are due to VMPC damage. In most circumstances, VMPC patients exhibit generally blunted affect and a specific defect of social emotions, but in response to direct personal frustration or provocation, VMPC patients may exhibit short-temper, irritability, and anger. In the moral judgement task we report here, participants respond to hypothetical actions and outcomes that elicit social emotions related to concern for others. In the Ultimatum Game, in contrast, participants respond to unfair take-it-or-leave-it offers that trigger frustration. In brief, the tasks in the two studies are different in that the Ultimatum Game involves self-interest in a real behavioural setting, whereas the task in the present study focuses on the interest of others described in a hypothetical scenario.

To conclude, the present findings are consistent with a model in which a combination of intuitive/affective and conscious/rational mechanisms operate to produce moral judgements[8,22,24–27]. Though the precise characterization of these potential systems awaits further work, the current results suggest that the VMPC is a critical neural substrate for the intuitive/affective but not for the conscious/rational system.

METHODS

Subjects. Six patients with bilateral, adult-onset damage to the VMPC and twelve brain-damaged comparison patients who had lesions that excluded structures thought to be important for emotions (VMPC, amygdala, insula, right somatosensory cortices) were recruited from the Patient Registry of the Division of Cognitive Neuroscience at the University of Iowa. Twelve healthy comparison subjects with no brain damage were recruited from the Iowa community. Groups were age-, gender- and ethnicity-matched. All participants gave written informed consent.

Neuroanatomical analysis. The neuroanatomical analysis of VMPC patients (Fig. 1) was based on magnetic resonance data for two subjects (those with lesions due to the surgical resection of orbital meningiomas) and on computerized tomography data for the other four subjects (with lesions due to rupture of an anterior communicating artery aneurysm). All neuroimaging data were obtained in the chronic epoch. Each patient's lesion was reconstructed in three dimensions using Brainvox[28]. Using the MAP-3 technique, the lesion contour for each patient was manually warped into a normal template brain. The overlap of lesions in this volume, calculated by the sum of n lesions overlapping on any single voxel, is colour-coded in Fig. 1.

Stimuli and task. Participants made judgements on a series of 50 hypothetical scenarios, which were adapted from a previously published set[8]. See the Supplementary Information for the full text of the actual scenarios used. Each scenario was presented as text through a series of three screens. The first two described the scenario and the third posed a question about a hypothetical action related to the scenario ("Would you … in order to …?"). Participants read and responded at their own pace, pressing an 'up' arrow key to advance from one

screen to the next, and a 'yes' or 'no' button to indicate an answer to the question. 'Yes' responses always indicated commission of the proposed action. There was no time limit for reading the scenario description (screens 1 and 2). Participants had a maximum of 25 s to read the final question screen and respond.

We used three classes of stimuli: non-moral scenarios ($n = 18$), and two classes of moral scenarios subdivided according to the emotional reaction elicited by the proposed action: 'personal' ($n = 21$) or 'impersonal' ($n = 11$), as described previously[7,8]. To validate this subdivision, an independent group of ten neurologically normal subjects rated the emotional salience of the actions proposed in the moral scenarios. The actions described in personal scenarios were rated as significantly more emotionally salient than the actions described in impersonal scenarios (means were 5.9 and 3.0 on a scale from 1 to 7, respectively; $t(31) = -8.90$, $P < 0.0001$). Within either class of moral scenarios (personal or impersonal), it was not valid to separately analyse judgements based on the emotional salience of the proposed action (that is 'high-emotion' versus 'low-emotion' scenarios) because emotionality ratings were remarkably similar for scenarios within each class: 9 of the 11 impersonal scenarios received a mean emotion rating between 1.1 and 3.0, while 20 of the 21 personal scenarios received a mean emotion rating between 5.3 and 6.7.

We further subdivided the personal moral scenarios into 'low-conflict' and 'high-conflict' on the basis of the reaction times and consensus produced on them by normal subjects. Reaction times on high-conflict scenarios were significantly longer than on low-conflict scenarios ($t(19) = -3.63$, $P = 0.002$). Importantly, low-conflict and high-conflict scenarios did not differ in their rated emotional salience ($t(19) = -0.85$, $P = 0.41$).

Received 3 November 2006; accepted 17 February 2007.
Published online 21 March 2007.

1. Eslinger, P. J., Grattan, L. M. & Damasio, A. R. Developmental consequences of childhood frontal lobe damage. *Arch. Neurol.* **49,** 764–769 (1992).
2. Anderson, S. W., Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neurosci.* **2,** 1032–1037 (1999).
3. Blair, R. J. R. A cognitive developmental approach to morality: investigating the psychopath. *Cognition* **57,** 1–29 (1995).
4. Mendez, M. F., Anderson, E. & Shapira, J. S. An investigation of moral judgment in frontotemporal dementia. *Cogn. Behav. Neurol.* **18,** 193–197 (2005).
5. Moll, J., de Oliveira-Souza, R., Bramati, I. E. & Grafman, J. Functional networks in emotional moral and nonmoral social judgments. *Neuroimage* **16,** 696–703 (2002).
6. Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H. P. & Villringer, A. An fMRI study of simple ethical decision-making. *Neuroreport* **14,** 1215–1219 (2003).
7. Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. An fMRI investigation of emotional engagement in moral judgment. *Science* **293,** 2105–2108 (2001).
8. Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. & Cohen, J. D. The neural bases of cognitive conflict and control in moral judgment. *Neuron* **44,** 389–400 (2004).
9. Luo, Q. *et al.* The neural basis of implicit moral attitude–An IAT study using event-related fMRI. *Neuroimage* **30,** 1449–1457 (2006).
10. Wheatley, T. & Haidt, J. Hypnotic disgust makes moral judgments more severe. *Psychol. Sci.* **16,** 780–784 (2005).
11. Valdesolo, P. & DeSteno, D. Manipulations of emotional context shape moral judgment. *Psychol. Sci.* **17,** 476–477 (2006).
12. Damasio, A. R., Tranel, D. & Damasio, H. Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behav. Brain Res.* **41,** 81–94 (1990).
13. Damasio, A. R. *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain* (Harcourt, New York, 2003).
14. Beer, J. S., Heerey, E. H., Keltner, D., Scabini, D. & Knight, R. T. The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *J. Pers. Soc. Psychol.* **85,** 594–604 (2003).
15. Kohlberg, L. *Essays on Moral Development* Vol. 1 *The Philosophy of Moral Development* (Harper Row, New York, 1981).
16. Damasio, A. R. *Descartes' Error: Emotion, Reason, and the Human Brain* (Penguin, New York, 1994).
17. Haidt, J. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol. Rev.* **108,** 814–834 (2001).
18. Ongur, D. & Price, J. L. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* **10,** 206–219 (2000).
19. Rolls, E. The orbitofrontal cortex and reward. *Cereb. Cortex* **3,** 284–294 (2000).
20. Koenigs, M. & Tranel, D. Irrational economic decision-making after ventromedial prefrontal damage: evidence from the ultimatum game. *J. Neurosci.* **27,** 951–956 (2007).
21. Anderson, S. W., Barrash, J., Bechara, A. & Tranel, D. Impairments of emotion and real-world complex behavior following childhood- or adult-onset damage to ventromedial prefrontal cortex. *J. Int. Neuropsychol. Soc.* **12,** 224–235 (2006).
22. Saver, J. L. & Damasio, A. R. Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia* **29,** 1241–1249 (1991).
23. Burgess, P. W. *et al.* The case for the development and use of "ecologically valid" measures of executive functions in experimental and clinical neuropsychology. *J. Int. Neuropsychol. Soc.* **12,** 194–209 (2006).
24. Hauser, M. D. *Moral Minds: How Nature Designed our Universal Sense of Right and Wrong* (Ecco/Harper Collins, New York, 2006).
25. Mikhail, J. *Rawls' Linguistic Analogy*. PhD thesis, Cornell Univ. (2000).
26. Cushman, F. A., Young, L. L. & Hauser, M. D. The role of conscious reasoning and intuition in moral judgments: Testing three principles of permissible harm. *Psychol. Sci.* **17,** 1082–1089 (2006).
27. Hauser, M. D., Cushman, F. A., Young, L. L., Jin, K-X. & Mikhail, J. A dissociation between moral judgments and justifications. *Mind Language* **22,** 1–21 (2006).
28. Frank, R. J., Damasio, H. & Grabowski, T. J. Brainvox: an interactive, multimodal visualization and analysis system for neuroanatomical imaging. *Neuroimage* **5,** 13–30 (1997).
29. Barrash, J. & Anderson, S. W. *The Iowa Rating Scales of Personality Change* (Department of Neurology, Univ. Iowa, Iowa, 1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Damasio for making available neuroanatomical analyses of lesion patients and for preparing Fig. 1. We thank all participants for their participation in the experiments and R. Saxe for comments on the manuscript. This work was supported by grants from the National Institutes of Health, the National Science Foundation, the Gordon and Betty Moore Foundation, and the Guggenheim Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.A. (radolphs@hss.caltech.edu).

ARTICLES

Opposing LSD1 complexes function in developmental gene activation and repression programmes

Jianxun Wang[1,2], Kathleen Scully[1]*, Xiaoyan Zhu[1]*, Ling Cai[1,3]*, Jie Zhang[1], Gratien G. Prefontaine[1], Anna Krones[1], Kenneth A. Ohgi[1], Ping Zhu[1], Ivan Garcia-Bassets[1], Forrest Liu[1], Havilah Taylor[1], Jean Lozach[4], Friederike L. Jayes[5], Kenneth S. Korach[5], Christopher K. Glass[4], Xiang-Dong Fu[4] & Michael G. Rosenfeld[1]

Precise control of transcriptional programmes underlying metazoan development is modulated by enzymatically active co-regulatory complexes, coupled with epigenetic strategies. One thing that remains unclear is how specific members of histone modification enzyme families, such as histone methyltransferases and demethylases, are used *in vivo* to simultaneously orchestrate distinct developmental gene activation and repression programmes. Here, we report that the histone lysine demethylase, LSD1—a component of the CoREST−CtBP co-repressor complex—is required for late cell-lineage determination and differentiation during pituitary organogenesis. LSD1 seems to act primarily on target gene activation programmes, as well as in gene repression programmes, on the basis of recruitment of distinct LSD1-containing co-activator or co-repressor complexes. LSD1-dependent gene repression programmes can be extended late in development with the induced expression of ZEB1, a Krüppel-like repressor that can act as a molecular beacon for recruitment of the LSD1-containing CoREST−CtBP co-repressor complex, causing repression of an additional cohort of genes, such as *Gh*, which previously required LSD1 for activation. These findings suggest that temporal patterns of expression of specific components of LSD1 complexes modulate gene regulatory programmes in many mammalian organs.

Epigenetic regulation by histone modification enzymes has important roles in the control of transcriptional programmes underlying metazoan development[1–3]. Investigation of NRSF/REST-mediated repression of a broad neurogenic programme[4,5] led to the discovery of the co-repressor CoREST[6] (also known as Rcor1), a SANT-domain-containing protein that interacts with specific histone deacetylases (HDACs)[7,8]. CoREST was subsequently found to be a component of a large complex including carboxy-terminal binding protein (CtBP) and the FAD-binding protein KIAA0601/BHC110/LSD1 (refs 9–14), which can function as both a histone diMe H3-K4 demethylase[15] and a diMe H3-K9 demethylase[16]. Recently, members of the large JmjC-domain-containing protein family (such as JHDMs and JMJDs) have been reported to function as specific histone H3-K9 or/and H3-K36 demethylases[17–19].

LSD1 is essential for mouse development

To begin investigating the roles of histone demethylases in development, we employed a strategy of conditional gene deletion of the *LSD1/KIAA0601* genomic locus by inserting *loxP* at sites flanking exon 6. This region corresponds to the amino terminus of the amine oxidase domain (Fig. 1a) and its deletion would result in an alteration of the reading frame of the *LSD1* (also known as *Aof2*) transcript. Using standard gene targeting technology in mouse embryonic stem cells, we obtained homologous recombination and generated both type I and type II recombinant alleles (Supplementary Fig. 1). Although mice heterozygous for conventional *LSD1* gene deletion (type I recombinant) appeared normal and fertile, no viable homozygous $LSD1^{-/-}$ embryos could be detected after embryonic day (E)7.5 (Supplementary Fig. 1). *LSD1* messenger RNA was essentially undetectable in $LSD1^{-/-}$ embryos and no truncated protein was detectable, indicating that the *LSD1* knockout strategy resulted in a null phenotype.

LSD1 controls pituitary terminal cell-type differentiation

To circumvent the early lethality of the $LSD1^{-/-}$ mice, we investigated the specific role of LSD1 during mammalian organogenesis by using the pituitary gland, in which *LSD1* is expressed throughout development, as a model system (Supplementary Fig. 2). Development of the anterior pituitary gland, in which five specific hormone-secreting cell types arise from a common primordium[20,21], has provided a powerful, well-defined model system for investigating the underlying epigenetic regulatory mechanisms that drive cell-type-specific gene expression programmes during organogenesis. We generated a pituitary-specific *LSD1* gene deletion using *Pitx1* Cre^{+} mice[22], which selectively express Cre recombinase in the oral ectoderm primordium of the pituitary and have been shown to execute effective recombination of the *ROSA26* locus in nearly all pituitary cells by E9–E9.5. The pituitary gland of the conditional *LSD1*-deleted embryo exhibited complete loss of *LSD1* transcript and protein starting from E9–E9.5 (Supplementary Fig. 2), but all the three lobes of the pituitary gland were maintained throughout development (Fig. 1b), suggesting that LSD1 is not required for morphogenesis of the pituitary gland. However, at E17.5, markers of terminal differentiation of *Pit1/Pou1f1*-lineage cells—growth hormone (*Gh*) and thyroid-stimulating hormone β (*Tshb*)—were

[1]Howard Hughes Medical Institute, Department and School of Medicine, University of California, San Diego, 9500 Gilman Drive, Room 345, La Jolla, California 92093-0648, USA. [2]Molecular Pathology Graduate Program, [3]Department of Biology Graduate Program, [4]Department of Cellular and Molecular Medicine, Department and School of Medicine, University of California, San Diego La Jolla, California 92093, USA. [5]National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA. *These authors contributed equally to this work.

undetectable in *LSD1*-deleted pituitaries, but were robustly expressed in pituitaries of littermate controls (Fig. 1c). Luteinizing hormone β (*Lhb*) and pro-opiomelanocortin (*Pomc1*), markers of the gonadotrope and corticotrope cell types, were also reproducibly decreased in *LSD1*-deleted pituitaries. Although residual POMC- and αGSU-encoding mRNAs were still present in *LSD1*-deleted pituitaries, very few if any POMC- or αGSU-positive cells could be detected by immunohistochemistry (Fig. 1d), suggesting that *LSD1* may regulate *POMC* and *αGSU* (also known as *Cga*) expression at both transcriptional and translational or/and protein stability levels. Therefore, LSD1 is essential for correct expression of all pituitary hormones.

The expression of *Pit1* and other cell-lineage-determining transcription factors, *Tbx19* and *Sf1*, was diminished, but remained easily detectable in the *LSD1*-deleted pituitaries (Fig. 2a). In addition, growth hormone releasing hormone receptor (*Ghrhr*) expression, a second marker of the somatotrope lineage[23–25], was undetectable in E17.5 *LSD1*-deleted pituitaries. Analyses of E12.5–E13.5 embryos revealed a normal morphology of the anterior pituitary gland, and normal expression of the initial cell-lineage determination factors, *Lhx3* and *Prop1*, in *LSD1*-deleted pituitaries (Fig. 2b). Thus, LSD1 controls late cell-lineage determination and terminal differentiation events but not early cell-lineage commitment events during pituitary development, indicating specificity of LSD1 in transcriptional control.

LSD1 developmentally regulates gene activation and repression

Despite a fivefold reduction in total *Pit1* transcripts, Pit1$^+$ cells, representing the somatotrope, thyrotrope and lactotrope precursors, were present at E17.5 in *LSD1*-deleted pituitaries, and Pit1 was robustly expressed in the positive cells (Fig. 3a). Intriguingly, Pit1, a critical transcription factor for *Gh* gene expression, was detected on the *Gh* promoter regulatory region by chromatin immunoprecipitation (ChIP) analysis of pituitary tissue as early as E13.5–E14.5, significantly before *Gh* mRNA is initially detected. However, recruitment of LSD1 to the same regulatory region of the *Gh* promoter occurred, only subsequently, on E16.5–E17.5 (Fig. 3b), coincident with *Gh* gene activation, suggesting that LSD1 is temporally and spatially positioned to exert critical roles in specific gene activation programmes that are required for terminal cell-type differentiation. Co-immunoprecipitation experiments using pituitary cells revealed Pit1–LSD1 interaction (Supplementary Fig. 3). In addition, LSD1 was recruited to the promoters of other Pit1-dependent genes, including prolactin (*Prl*) and *Tshb*, as well as *Pit1*, as determined by ChIP analysis, with similar results using either of two independently derived, specific anti-LSD1 antibodies (Fig. 3c). Therefore, LSD1 seems to regulate gene activation programmes during pituitary organogenesis, consistent with previous findings that LSD1 is required for gene activation[16,18].

During pituitary development, LSD1 seems to regulate, directly or indirectly, gene repression programmes as well. For example, RNA profiling analysis and *in situ* hybridization revealed markedly increased expression of *Hey1*, a direct downstream target of the Notch signalling

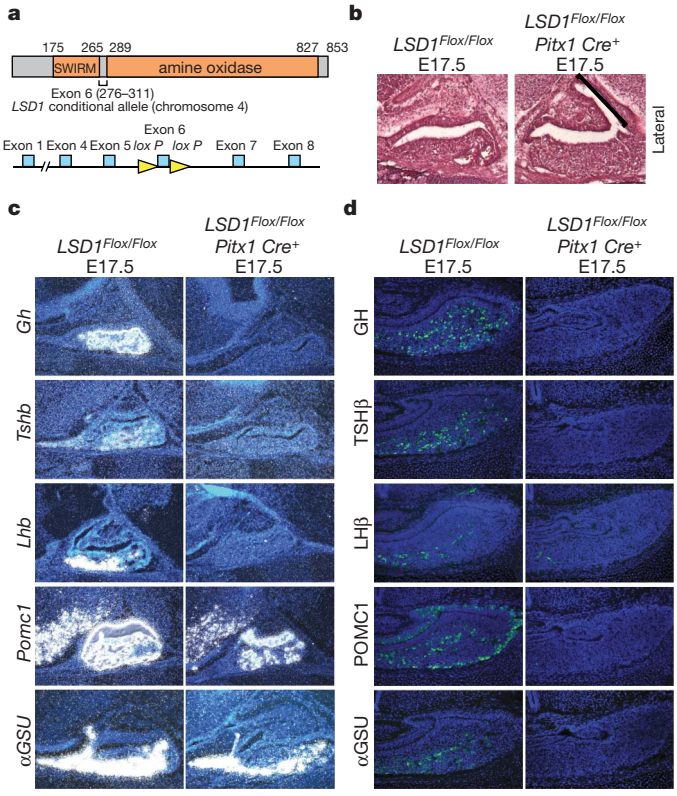


Figure 1 | LSD1 is required for cell-type-specific pituitary hormone expression. **a**, The strategy for generation of a conditional allele of *LSD1*. **b**, Haematoxylin and eosin staining of control versus *LSD1*-deleted pituitaries; a representative lateral para-sagittal section of E17.5 pituitaries is shown. **c**, Gene expression of *Gh*, *Tshb*, *Lhb*, *Pomc1* and *αGSU/Cga* in control or *LSD1*-deleted pituitaries at E17.5 by *in situ* hybridization; a representative sagittal section of pituitary gland is shown. **d**, Protein expression of GH, TSHβ, LHβ, POMC (expression was detected using αACTH immunoglobulin G) and αGSU in control or *LSD1*-deleted pituitaries at E17.5 by immunohistochemistry; a representative coronal section of pituitary gland shown is shown.

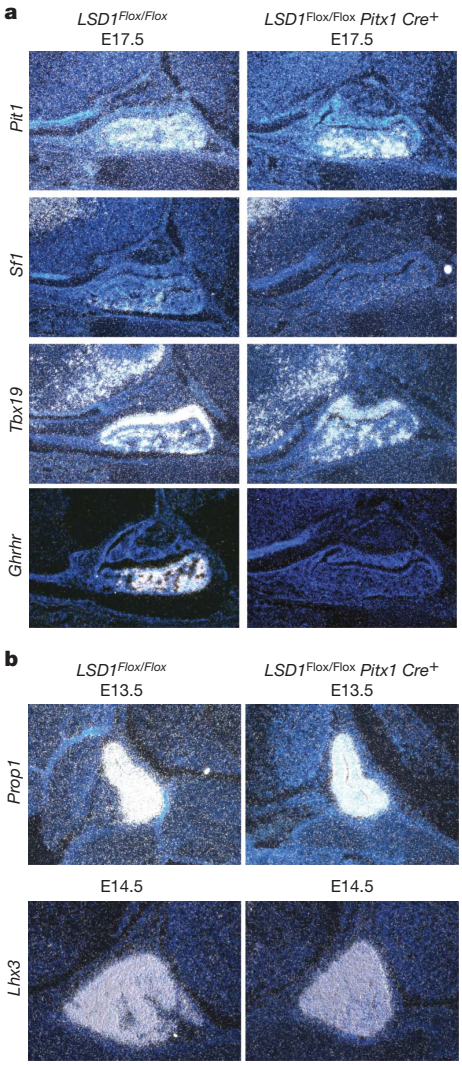


Figure 2 | LSD1 is involved in regulatory programmes in pituitary development. **a**, Expression of *Pit1*, *Sf1*, *Tbx19* and *Ghrhr* in control or *LSD1*-deleted pituitaries at E17.5 by *in situ* hybridization. **b**, Gene expression of *Prop1* and *Lhx3* in control or *LSD1*-deleted pituitaries at E13.5–E14.5 by *in situ* hybridization.

pathway[26] in the E17.5 *LSD1*-deleted pituitary, indicating that LSD1 may also have a critical role in CtBP-mediated repression of Notch signalling pathways[27] (Fig. 3d). Co-immunoprecipitation experiments revealed an interaction between the Notch-interacting transcription factor RBP-J/CSL/RBPSUH and LSD1 (Fig. 3e), consistent with known interactions between CSL and CtBP, and ChIP analysis of E17.5 pituitary revealed recruitment of LSD1 to the *Hey1* promoter (Fig. 3f). In addition, LSD1 functions as a repressor for several other transcription units as well (Supplementary Fig. 4), including several cell cycle control genes, suggesting potential roles of LSD1 on growth control in developing pituitary. Examination of cell proliferation in *LSD1*-deleted pituitaries revealed an increased number of luminal cells in S phase, beginning on E13.5 and based on the results of two-hour BrdU pulse-labelling experiments[28] (Supplementary Fig. 3), with no evidence of altered apoptosis (data not shown). *LSD1*-deleted pituitaries exhibited some dysmorphogenesis—a convoluted lumen at E17.5 resembling that of *Prop1*-deleted pituitaries[22,29–31]—and exhibited more Ki67-positive luminal cells (Supplementary Fig. 3), indicating increased proliferation. Among LSD1-regulated cell cycle control genes revealed by RNA profiling analysis, we confirmed the upregulation of cyclin E1 (*Ccne1*) and *Id2* at E17.5 (Supplementary Figs 3 and 4). ChIP revealed that LSD1, CoREST and CtBP were recruited to the *Ccne1* promoter in the E17.5 pituitary (Fig. 3g), consistent with recruitment of the CtBP–LSD1-containing co-repressor complex[9]. Furthermore, there was a small, but highly reproducible increase in histone diMe H3-K4 on the *Ccne1* promoter at E17.5 in *LSD1*-deleted pituitaries as measured by quantitative PCR (qPCR) (data not shown). Therefore, *Ccne1* seems to be a direct, negatively regulated target of LSD1 *in vivo*. Consistent with retinoblastoma (*Rb*) repression of *Ccne1* gene expression[32] and roles of *Id2* in pituitary tumour progression in $Rb^{+/-}$ mice[33,34], we found that *Cdc2*- and *Ccne1*-promoter-driven reporters were inhibited by re-expression of wild-type, but not enzymatically inactive, LSD1 in $LSD1^{-/-}$ mouse embryonic fibroblast cell lines (MEFs) generated from $LSD1^{Flox/Flox}$ MEFs (Fig. 3h).

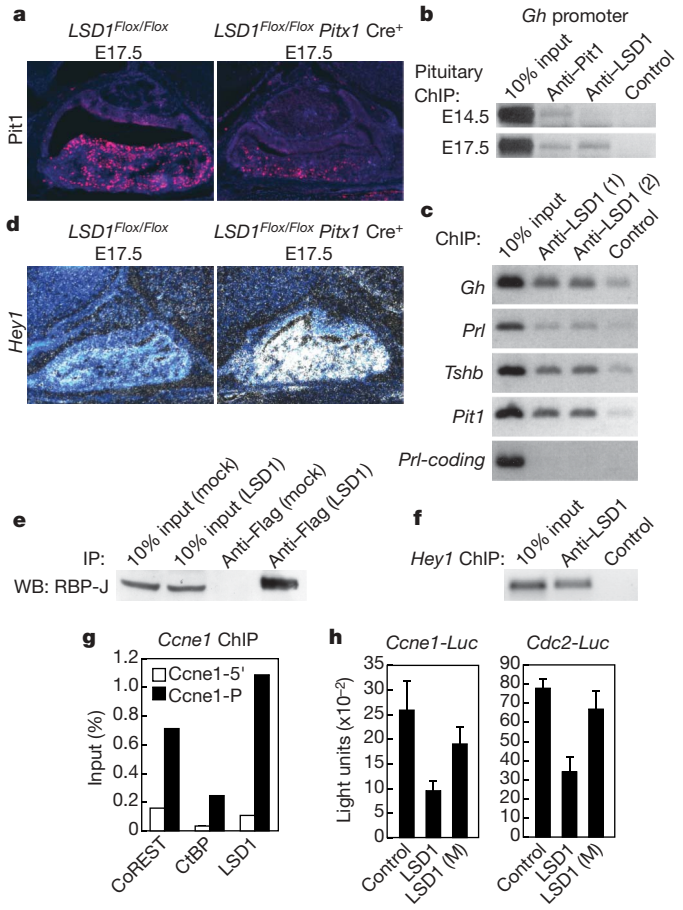


Figure 3 | LSD1 regulates gene expression during pituitary development. **a**, Immunohistochemical analysis of Pit1 lineage precursors in control or *LSD1*-deleted pituitaries at E17.5 using anti-Pit1 immunoglobulin G. **b**, Chromatin immunoprecipitation (ChIP) with anti-Pit1 and anti-LSD1 immunoglobulin G to analyse their occupancy on the *Gh* promoter in E13.5–E14.5 or E16.5–E17.5 pituitaries. **c**, A ChIP assay was performed on P10 pituitaries using two independent, specific anti-LSD1 immunoglobulin G antibodies on the *Gh*, *Prl*, *Tshb* and *Pit1* promoters with the *Prl* coding region providing a negative control. **d**, Expression of *Hey1* transcripts in control or *LSD1*-deleted pituitaries at E17.5 by *in situ* hybridization. **e**, Co-immunoprecipitation assay performed with HEK293 cells using Flag-tagged, transfected LSD1 and analysed with anti-RBP-J immunoglobulin G by western blot (WB). **f**, ChIP with the *Hey1* promoter in E17.5 pituitary with anti-LSD1 or pre-immune control immunoglobulin G. **g**, ChIP with the *Ccne1* promoter (*Ccne1-P*) and 5′ upstream control region (*Ccne1-5′*), using anti-CtBP, anti-CoREST or anti-LSD1 immunoglobulin G. **h**, Wild-type LSD1 represses *Cdc2* and *Ccne1* promoters in a reporter assay in $LSD1^{-/-}$ MEF cells. Error bars, s.e.m.

LSD1–ZEB1 complex regulates GH gene repression in lactotropes

On the basis of the diversity of LSD1-containing complexes, we wished to assess whether the role of LSD1 on a given target gene might itself be altered during development, as a regulatory aspect of cell-type determination. In the pituitary, a final cell-type differentiation (appearance of GH^-/PRL^+ lactotropes, beginning at post-partum (P) P5–P10) has been previously suggested to reflect a 'switch' of GH^+/PRL^- cells to GH^-/PRL^+ cells[35,36], but it can be alternatively proposed that lactotropes arise independently after birth from cells that have not previously expressed *Gh*. In either model, the molecular events underlying *Gh* gene restriction in emerging lactotropes have remained unknown. Previous *in vivo* experiments suggested that a sequence (−161/−146) in the rat *Gh* promoter was a critical repressive site required for *Gh* gene restriction events in developing lactotropes[37]. Screening of a pituitary complementary DNA phage expression library, using the −161/−146 region as probe, identified cDNA expression clones encoding three independent, overlapped C-terminal portions of ZEB1 (also known as Tcf8/δEF1/BZP/zfhx1a/zfhep/AREB6), a 130 kDa protein containing one homeodomain and seven Cys/His zinc fingers[38]. ZEB1 contains three PLDLS CtBP recognition sequences[39], consistent with it being identified as a component of the LSD1–CoREST–CtBP complex isolated from HeLa cells[9] and with it exerting a role in repression. Interestingly, a thyroid hormone receptor response element (T_3RE) is located immediately 5′ to the ZEB1 response element (ZEB1RE) and functions both in *Gh* gene activation in somatotropes and repression in lactotropes[37]. LCoR/KIAA1795/MLR2, another component of the ZEB1-containing CtBP complex and first identified as an agonist-dependent oestrogen receptor co-repressor[40], can function as a repressor on the T_3R-regulated promoter of *Dio1*, in the presence of T_3 (Supplementary Fig. 5). Its repression function depends on both its LXXLL and PLDLS motifs, indicating that LCoR might also contribute to repression of *Gh* because of interactions with both the T_3R (through its LXXLL) and ZEB1–LSD1–CoREST–CtBP complex (through its PLDLS), perhaps by inhibiting the normal ligand-dependent cycling of co-activators on T_3R[41].

On the basis of the CtBP–LSD1 complex isolated from HeLa cells[9], we expected continuous, coordinate expression of all components throughout pituitary development. However, examination of the ontogeny of *Zeb1* revealed only minimal expression in the pituitary before birth, but a marked induction beginning at P5–P10 (Fig. 4a), a temporal pattern consistent with a potential role for ZEB1 in *Gh* gene restriction events. The same pattern of expression proved to be the case for two additional components of the LSD1–CoREST–CtBP complex, LCoR and a polycomb co-repressor PC2 (also known as CBX4; ref. 42) (Fig. 4b, c), which functions as a SUMO-E3 ligase[43]. Hence, several components present in the CtBP–LSD1-containing co-repressor complex, initially co-isolated in HeLa cells, actually seem to exhibit temporal patterns of expression during the postpartum period of pituitary development.

Co-immunoprecipitation and ChIP analyses were performed to confirm the presence of the ZEB1−CtBP−CoREST−LSD1 complex in the pituitary cells (Supplementary Fig. 6) and that ZEB1, LSD1 and LCoR were all present, along with Pit1, on the *Gh* promoter in the P20 pituitary (Fig. 4d).

Because ZEB1 can bind the *Gh* promoter major repressive element that is responsible, *in vivo*, for late *Gh* restriction in the emerging lactotrope cell population[37], we next examined whether ZEB1 was indeed bound only in cell types in which *Gh* is repressed (that is, lactotropes), or whether it was also bound to the *Gh* promoter in *Gh*-expressing cell types (that is, somatotropes) in the postpartum pituitary. Because it is currently not possible to obtain significant numbers of pure somatotropes and lactotropes, we took advantage of previous observations from genome-wide promoter analyses, indicating that the triMe H3-K4 histone mark was invariantly found on active, NCoR-negative promoters. Therefore, a two-step ChIP analysis was performed using either anti-triMe-H3-K4 or anti-NCoR immunoglobulin G in the first round, and specific anti-ZEB1 and control immunoglobulin G in the second round. We found that ZEB1 was not detected on anti-triMe-H3-K4-positive, active promoters, but was invariantly present on anti-NCoR-positive, repressed *Gh* promoters (Fig. 4e), as occurs in lactotropes[37]. ChIP analysis performed using E17.5 and adult pituitaries revealed that detectable recruitment of ZEB1 to the *Gh* promoter was noted only after birth (Fig. 4f).

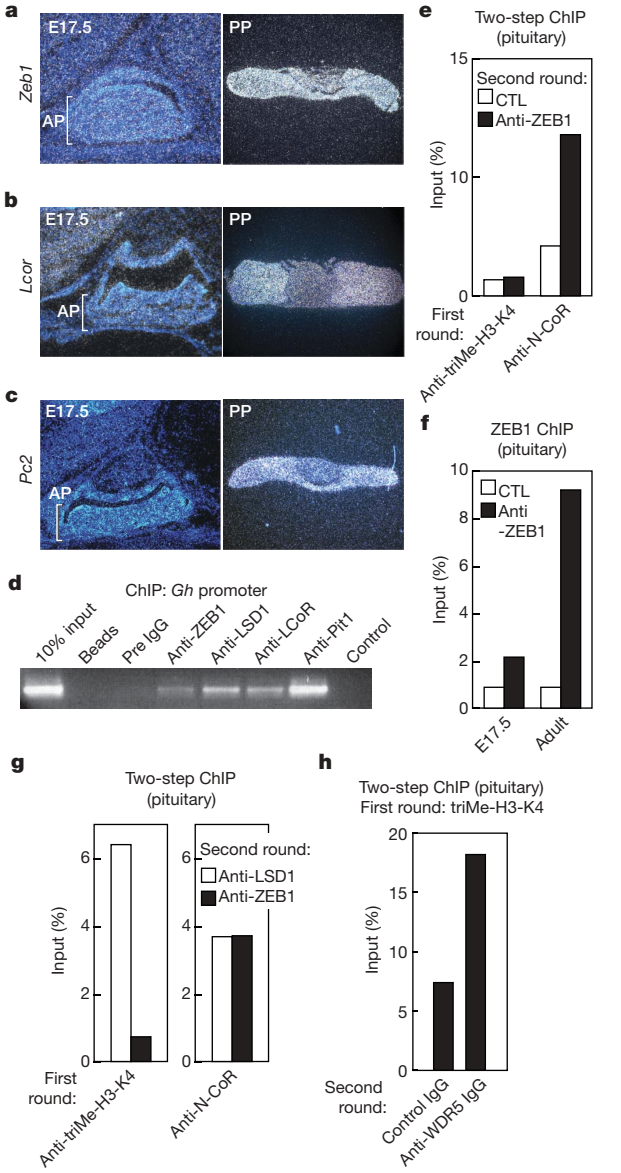


Figure 4 | Postpartum induction of the LSD1 complex in *Gh* restriction events. **a**, Ontogeny of *Zeb1* expression in pituitary development. AP, anterior pituitary. **b**, Ontogeny of *Lcor* expression in pituitary development. **c**, Ontogeny of *Pc2* expression in pituitary development. In **a**–**c**, the postpartum pituitary (PP) was photographed at lower magnification (×2.5), so it could be shown in a single field. **d**, ZEB1/LSD1/LCoR/Pit1 were recruited to the *Gh* promoter by ChIP assays performed using P20 pituitaries. **e**, Two-step ChIP analysis in adult pituitary was performed using anti-triMe-H3-K4 or anti-NCoR immunoglobulin G in the first round, and anti-ZEB1 and control immunoglobulin G in the second round. **f**, ChIP analysis was performed to assess the recruitment of ZEB1 on the *Gh* promoter in embryonic (E17.5) and young adult (P20) pituitaries. **g**, ChIP analysis of LSD1 and ZEB1 occupancy on the active and repressed *Gh* promoters using two-step ChIP analysis. **h**, Two-step ChIP analysis with the *Gh* promoter in young adult (P20) pituitary using anti-triMe-H3-K4 immunoglobulin G in the first round and anti-WDR5 and control immunoglobulin G in the second round.

Consistent with our *in vivo* data that LSD1 is required for *Gh* gene activation, we found that LSD1 was present both on triMe-H3-K4$^+$ *Gh* promoters and on repressed, NCoR$^+$ *Gh* promoters (Fig. 4g). On the triMe H3-K4$^+$ *Gh* promoters, two-step ChIP also revealed the co-recruitment of WDR5, a canonical component of the MLL1 complex[12,44] (Fig. 4h). Together, these findings are consistent with recruitment of LSD1 in two distinct complexes to the *Gh* promoter, exerting a role in gene activation as a component of the MLL1 co-activator complex in the absence of ZEB1 binding, and exerting a role in *Gh* repression in lactotropes, as part of the ZEB1−CoREST−CtBP−LSD1 co-repressor complex.

Consistently, using two pituitary cell lines, MMQ and GC, which expressed *Prl* or *Gh* respectively[45], we found an enhanced recruitment of CtBP, LCoR, HDAC1, HDAC2 and PC2, as well as NCoR, on the *Gh* promoter in GH$^-$/PRL$^+$ MMQ cells, compared with GH$^+$ GC pituitary cells (Supplementary Fig. 7). The levels of diMe H3-K4 are higher on the *Gh* promoter in GC cells than in the MMQ cells, whereas levels of diMe H3-K9 are higher in MMQ cells than in GC cells. Further, using MMQ cells, even with a poor transfection efficiency, we found that some derepression of *Gh* could be achieved by *Zeb1* short interfering RNA (siRNA) and siRNAs against several other components of the LSD1−CoREST−CtBP complex, but not, for example, by *Zeb2* siRNA (Supplementary Fig. 7). In addition, siRNA-mediated knockdown of the histone H3-K9 methyltransferases EuHMT1 and EuHMT2/G9a, also found in the CtBP complex[9], partially derepresses *Gh* gene expression in MMQ lactotropes, suggesting that histone H3-K9 methylation also contributes to the repression events. Together, these data suggest a potential model in which ZEB1, expressed in the pituitary after birth, acts as a molecular beacon that causes recruitment of an LSD1−CoREST−CtBP−LCoR-containing complex to a cognate repressive element on the *Gh* promoter, thereby restricting *Gh* expression out of the developing lactotrope. Hence, the role of ZEB1 is conceptually similar to the role of REST in LSD1−CoREST complex-mediated repression[6,15].

Although the precise mechanism(s) that induce postpartum expression of *Zeb1*, *Lcor* and *Pc2* clearly remain incompletely determined, the quantitative expression of these regulators is increased by oestrogen (E_2). Indeed, in *ERα*/*Esr1*-deleted mice[46], levels of *Zeb1*, *Pc2* and *Lcor* are at least 2−3-fold diminished in pituitaries compared with littermate controls (Fig. 5a and data not shown), indicating that E_2 is one, but clearly not the sole, signal resulting in expression of these modules that target the LSD1 co-repressor complex to the *Gh* gene; this is consistent with an observation of oestrogen induction of *Zeb1* gene expression in chicken oviduct[47]. Even in heterologous cell lines, such as U2OS (ERα$^+$/ERβ$^+$) cells, E_2 also causes induction of *Zeb1*, *Pc2* and *Lcor* (Supplementary Fig. 5), and in the presence of Pit1, E_2 caused a striking repression of *Gh* reporter expression, in contrast to a simultaneous induction of *Prl* reporter expression (Fig. 5b). This system allowed us to test the relative roles of components of the co-repressor

complex in repression using specific siRNAs to achieve effective depletion of each mRNA in these cells (Supplementary Fig. 5). Robust derepression of a *Gh*-driven reporter could be observed with siRNAs against *Zeb1, LSD1, Lcor, Ctbp* and *Pc2*, each causing a partial relief of the repressed state (Fig. 5c). These data are consistent with the roles of E_2 in the restriction events, which is correlated with the initial appearance of prolactin-producing cells in concert with a postpartum oestradiol-17β (E_2) surge[48,49].

How might this cell-type-selective recruitment of ZEB1 be regulated? Intriguingly, both ZEB1 and ZEB2 have been reported to be sumoylated on two conserved lysine residues[50], with one residue (ZEB1 K697) located adjacent to a CtBP interaction motif of ZEB1. The SUMO-E3 ligase PC2 induced in postpartum pituitary can use ZEB1 as a substrate[50]. Using the *ZEB1RE* region of the *Gh* promoter to drive a reporter in HEK293T cells, we found that wild-type ZEB1 causes repression, but mutation of these known sumoylated lysine residues (K271R, K697R) actually caused a loss of the repression function of ZEB1 (Fig. 5d), indicating that sumoylation of ZEB1 may be required for its full repression function. To test for the possible functional significance of this sumoylation, we tested for the ability of ZEB1 to bind to its cognate site on the *Gh* gene by using HA-tagged, wild-type or mutant ZEB1 (K271R, K697R) and evaluating by ChIP the binding to *Gh ZEB1RE* in the co-transfected reporter plasmid. Surprisingly, the ability of ZEB1 to be recruited proved to be highly dependent on the presence of the *ZEB1RE* and its ability to be sumoylated (Fig. 5e). Further, we found that recruitment of LSD1 was dependent on the presence of the *ZEB1RE* (Fig. 5f). Intriguingly, we found that treatment with E_2 also caused increased levels of sumoylated ZEB1 (Fig. 5g), consistent with transcriptional and post-transcriptional roles of E_2 levels in mediating restriction of the *Gh* transcription unit. Expression of PC2 is highly enriched in MMQ/lactotrope versus GC/somatotrope cell types (Supplementary Fig. 5), again indicating that ZEB1 sumolyation might exert a cell-type-specific role.

Conclusions

We have thus demonstrated that LSD1 regulates specific developmental programmes in the window following initial organ commitment and before cell-type differentiation, with LSD1 serving initially as a key component of opposing co-activator and co-repressor complexes that are recruited in a gene-specific fashion that is required for dictating specific programmes of gene expression during mammalian organogenesis (Fig. 5h and Supplementary Fig. 8). We consider it likely that LSD1 activates a cohort of gene targets, including *Gh* expression, by removal of repressive histone marks[16,18], but non-histone substrates are likely to be biologically important. A second, intriguing aspect of regulation of *Gh* expression is the temporally regulated, cell-type-specific, restriction of *Gh* expression in emerging lactotropes after birth, apparently due to, at least in part, postpartum induction of *ZEB1*, a Krüppel-like zinc finger DNA-binding repressor, and at least two other components of the CtBP−CoREST−LSD1 co-repressor complex, together assembling an LSD1-containing co-repressor complex on the *Gh* promoter. Thus, after birth, LSD1 is present on the *Gh* promoter in a co-activator

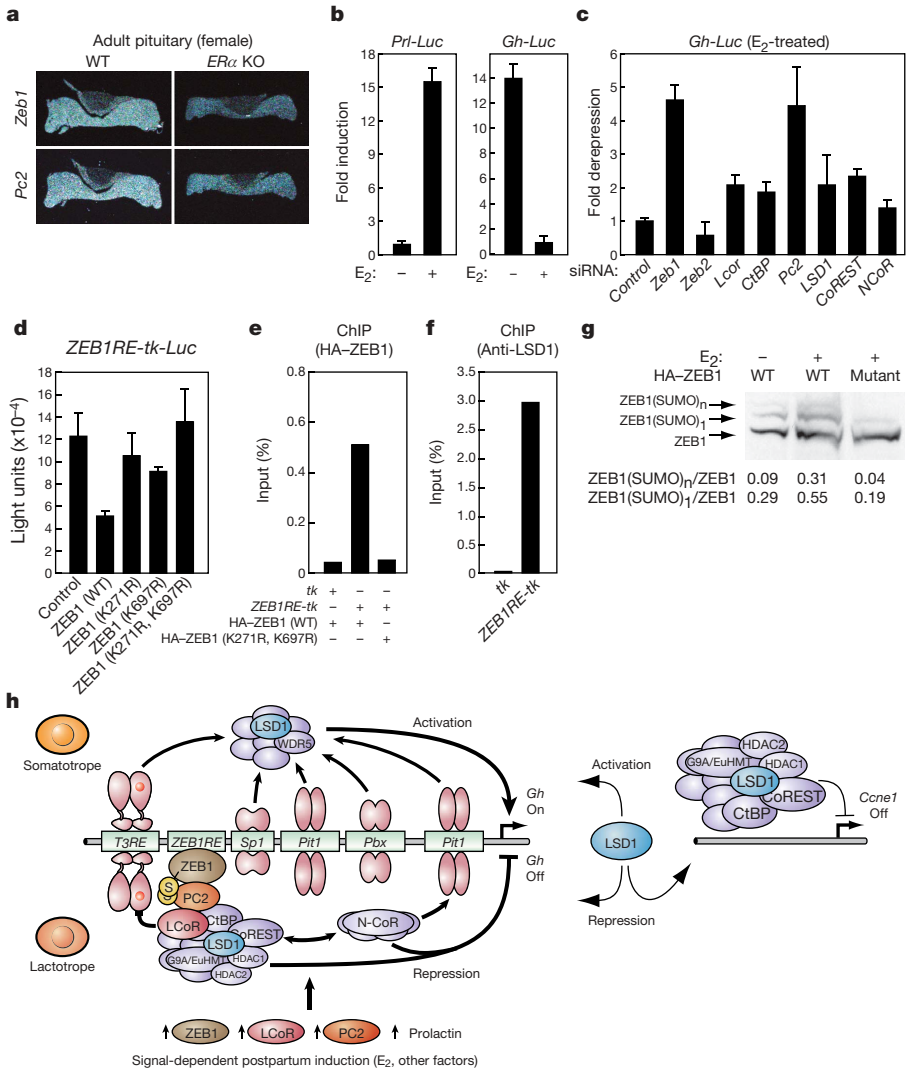


Figure 5 | The LSD1 complex has roles in restriction of *Gh* expression in lactotropes. **a**, Analysis of *Zeb1* and *Pc2* expression in adult female pituitaries of *ERα* knockout (KO) mice versus their wild-type littermates by *in situ* hybridization. **b**, Repression of *Gh-Luc* reporter and induction of *Prl-Luc* reporters in response to E_2 in U2OS ($ER\alpha^+/ER\beta^+$) cells. **c**, Specific siRNA against individual components of the ZEB1−CtBP−CoREST−LSD1 complex were assessed for the ability to reverse the repression by assaying *Gh-Luc* reporter expression. **d**, ZEB1 represses expression of the *Gh ZEB1RE*-thymidine kinase *(tk)-Luc* reporter in a SUMO-dependent fashion. **e**, ChIP analysis revealed that ZEB1 recruitment was dependent on the *Gh ZEB1RE* site and 'SUMOylation' of ZEB1. **f**, Recruitment of LSD1 was assessed in transient transfection assays in HEK293T cells using control *tk-Luc* or *ZEB1RE/tk-Luc* reporters, revealing that binding requires the presence of the *ZEB1RE* site. **g**, Induction of ZEB1 mono-sumoylation ($ZEB1(SUMO)_1$) or multiple-sumoylation ($ZEB1(SUMO)_n$) in response to oestrogen in U2OS ($ER\alpha^+/ER\beta^+$) cells; the ratio of sumoylated versus unmodified ZEB1 was quantified by NIH imageJ software. **h**, Model of the roles of distinct LSD1-containing complexes during pituitary organogenesis.

complex in somatotropes, and in a ZEB1-recruited co-repressor complex in lactotropes.

METHODS

Generation and genotyping of LSD1-deficient mice. *LSD1* conditional allele mice were generated by targeted mutagenesis in embryonic stem cells to insert two *loxP* sites, flanking exon 6 of *LSD1* (Supplementary Fig. 1); correct targeting was established by Southern blotting with 5′ and 3′ external probes. Embryos were genotyped using PCR. *Pitx1 Cre*$^{+}$ mice were generated as described previously[22]. *LSD1*$^{Flox/Flox}$ mice were crossed with *Pitx1 Cre*$^{+}$ lines. Pituitary-specific *LSD1*-deleted mice were generated through breeding *LSD1*$^{Flox/Flox}$ mice with *LSD1*$^{Flox/+}$ / *Pitx1 Cre*$^{+}$ mice.

***In situ* hybridization and immunohistochemistry.** *In situ* hybridization was performed as described previously[22]. For immunohistochemistry, 10% neutral-buffer-formalin-fixed, 20% sucrose-penetrated embryos were embedded with paraffin and sectioned with a thickness of 12 μm for histology slides. Unstained sections were post-fixed with 10% neutral buffer formalin for 10 min and washed twice with PBS. Antigen was retrieved through boiling in 10 mM citrate buffer (pH 6.0) for 10 min. Immunostaining was carried out with standard immunohistochemistry protocols.

Received 1 December 2006; accepted 5 February 2007.
Published online 28 March 2007.

1. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression, how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33**, (Suppl.) 245–254 (2003).
2. Fischle, W., Wang, Y. & Allis, C. D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425,** 475–479 (2003).
3. Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nature Rev. Mol. Cell Biol.* **6,** 838–849 (2005).
4. Chong, J. A. *et al.* REST, a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80,** 949–957 (1995).
5. Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF), a coordinate repressor of multiple neuron-specific genes. *Science* **267,** 1360–1363 (1995).
6. Andres, M. E. *et al.* CoREST, a functional corepressor required for regulation of neural-specific gene expression. *Proc. Natl Acad. Sci. USA* **96,** 9873–9878 (1999).
7. Ballas, N. *et al.* Regulation of neuronal traits by a novel transcriptional complex. *Neuron* **31,** 353–365 (2001).
8. Lunyak, V. V. *et al.* Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* **298,** 1747–1752 (2002).
9. Shi, Y. *et al.* Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422,** 735–738 (2003).
10. You, A., Tong, J. K., Grozinger, C. M. & Schreiber, S. L. CoREST is an integral component of the CoREST–human histone deacetylase complex. *Proc. Natl Acad. Sci. USA* **98,** 1454–1458 (2001).
11. Humphrey, G. W. *et al.* Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *J. Biol. Chem.* **276,** 6817–6824 (2001).
12. Nakamura, T. *et al.* ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol. Cell* **10,** 1119–1128 (2002).
13. Hakimi, M. A. *et al.* A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc. Natl Acad. Sci. USA* **99,** 7420–7425 (2002).
14. Hakimi, M. A. *et al.* A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. *J. Biol. Chem.* **278,** 7234–7239 (2003).
15. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119,** 941–953 (2004).
16. Metzger, E. *et al.* LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437,** 436–439 (2005).
17. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439,** 811–816 (2005).
18. Yamane, K. *et al.* JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125,** 483–495 (2006).
19. Whetstine, J. R. *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125,** 467–481 (2006).
20. Keegan, C. E. & Camper, S. A. Mouse knockout solves endocrine puzzle and promotes new pituitary lineage model. *Genes Dev.* **17,** 677–682 (2003).
21. Scully, K. M. & Rosenfeld, M. G. Pituitary development, regulatory codes in mammalian organogenesis. *Science* **295,** 2231–2235 (2002).
22. Olson, L. E. *et al.* A homeodomain-mediated mechanism for β-catenin-dependent switching events dictates cell lineage determination. *Cell* **125,** 593–605 (2006).
23. Mayo, K. E. Molecular cloning and expression of a pituitary-specific receptor for growth hormone-releasing hormone. *Mol. Endocrinol.* **6,** 1734–1744 (1992).
24. Lin, C., Lin, S. C., Chang, C. P. & Rosenfeld, M. G. Pit-1-dependent expression of the receptor for growth hormone releasing factor mediates pituitary cell growth. *Nature* **360,** 765–768 (1992).
25. Godfrey, P. *et al.* GHRH receptor of *little* mice contains a missense mutation in the extracellular domain that disrupts receptor function. *Nature Genet.* **4,** 227–232 (1993).
26. Maier, M. M. & Gessler, M. Comparative analysis of the human and mouse *Hey1* promoter: *Hey* genes are new Notch target genes. *Biochem. Biophys. Res. Commun.* **275,** 652–660 (2000).
27. Oswald, F. *et al.* RBP-Jκ/SHARP recruits CtIP/CtBP corepressors to silence Notch target genes. *Mol. Cell. Biol.* **25,** 10379–10390 (2005).
28. Zhu, X. *et al.* Sustained Notch signaling in progenitors is required for sequential emergence of distinct cell lineages during organogenesis. *Genes Dev.* **20,** 2739–2753 (2006).
29. Sornson, M. W. *et al.* Pituitary lineage determination by the Prophet of *Pit-1* homeodomain factor defective in Ames dwarfism. *Nature* **384,** 327–333 (1996).
30. Ward, R. D. *et al.* Role of PROP1 in pituitary gland growth. *Mol. Endocrinol.* **19,** 698–710 (2005).
31. Gage, P. J., Roller, M. L., Saunders, T. L., Scarlett, L. M. & Camper, S. A. Anterior pituitary cells defective in the cell-autonomous factor, df, undergo cell lineage specification but not expansion. *Development* **122,** 151–160 (1996).
32. Botz, J. *et al.* Cell cycle regulation of the murine cyclin E gene depends on an E2F binding site in the promoter. *Mol. Cell. Biol.* **16,** 3401–3409 (1996).
33. Hu, N. *et al.* Heterozygous *Rb-1 delta 20/+* mice are predisposed to tumors of the pituitary gland with a nearly complete penetrance. *Oncogene* **9,** 1021–1027 (1994).
34. Lasorella, A., Rothschild, G., Yokota, Y., Russell, R. G. & Iavarone, A. Id2 mediates tumor initiation, proliferation, and angiogenesis in *Rb* mutant mice. *Mol. Cell. Biol.* **25,** 3563–3574 (2005).
35. Behringer, R. R., Mathews, L. S., Palmiter, R. D. & Brinster, R. L. Dwarf mice produced by genetic ablation of growth hormone-expressing cells. *Genes Dev.* **2,** 453–461 (1988).
36. Borrelli, E., Heyman, R. A., Arias, C., Sawchenko, P. E. & Evans, R. M. Transgenic mice with inducible dwarfism. *Nature* **339,** 538–541 (1989).
37. Scully, K. M. *et al.* Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science* **290,** 1127–1131 (2000).
38. Williams, T. M. *et al.* Identification of a zinc finger protein that inhibits *IL-2* gene expression. *Science* **254,** 1791–1794 (1991).
39. Postigo, A. A. & Dean, D. C. ZEB represses transcription through interaction with the corepressor CtBP. *Proc. Natl Acad. Sci. USA* **96,** 6683–6688 (1999).
40. Fernandes, I. *et al.* Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and -independent mechanisms. *Mol. Cell* **11,** 139–150 (2003).
41. Perissi, V. & Rosenfeld, M. G. Controlling nuclear receptors, the circular logic of cofactor cycles. *Nature Rev. Mol. Cell Biol.* **6,** 542–554 (2005).
42. Satijn, D. P. *et al.* Interference with the expression of a novel human polycomb protein, hPc2, results in cellular transformation and apoptosis. *Mol. Cell. Biol.* **17,** 6076–6086 (1997).
43. Kagey, M. H., Melhuish, T. A. & Wotton, D. The polycomb protein Pc2 is a SUMO E3. *Cell* **113,** 127–137 (2003).
44. Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121,** 859–872 (2005).
45. Ooi, G. T., Tawadros, N. & Escalona, R. M. Pituitary cell lines and their endocrine applications. *Mol. Cell. Endocrinol.* **228,** 1–21 (2004).
46. Korach, K. S. Insights from the study of animals lacking functional estrogen receptor. *Science* **266,** 1524–1527 (1994).
47. Chamberlain, E. M. & Sanders, M. M. Identification of the novel player δEF1 in estrogen transcriptional cascades. *Mol. Cell. Biol.* **19,** 3600–3606 (1999).
48. Lieberman, M. E., Slabaugh, M. B., Rutledge, J. J. & Gorski, J. The role of estrogen in the differentiation of prolactin producing cells. *J. Steroid Biochem.* **19,** 275–281 (1983).
49. Day, R. N., Koike, S., Sakai, M., Muramatsu, M. & Maurer, R. A. Both Pit-1 and the estrogen receptor are required for estrogen responsiveness of the rat prolactin gene. *Mol. Endocrinol.* **4,** 1964–1971 (1990).
50. Long, J., Zuo, D. & Park, M. Pc2-mediated sumoylation of Smad-interacting protein 1 attenuates transcriptional repression of E-cadherin. *J. Biol. Chem.* **280,** 35477–35489 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Nelson and L.Wang for technical assistance, L. Olson for generating *Pitx1 Cre*$^{+}$ mice, Y. Geng for the *Ccne1* promoter reporter, W. Herr for the generous gift of anti-WDR5 antibody, A. Gonzalez, J. Liu for advice on reagents, and J. Hightower and M. Fisher for figure and manuscript preparation. M.G.R is an investigator with HHMI. This work was supported by grants from the NIH and NCI to M.G.R., C.K.G. and X.-D.F.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.G.R. (mrosenfeld@ucsd.edu).

Vol 446|19 April 2007|doi:10.1038/nature05676

LETTERS

Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements

Liana F. Lareau[1]*, Maki Inada[2]*, Richard E. Green[1]†, Jordan C. Wengrod[1] & Steven E. Brenner[1,2]

The human and mouse genomes share a number of long, perfectly conserved nucleotide sequences, termed ultraconserved elements[1]. Whereas these regions can act as transcriptional enhancers when upstream of genes, those within genes are less well understood. In particular, the function of ultraconserved elements that overlap alternatively spliced exons of genes encoding RNA-binding proteins is unknown[1,2]. Here we report that in every member of the human SR family of splicing regulators, highly or ultraconserved elements are alternatively spliced, either as alternative 'poison cassette exons' containing early in-frame stop codons, or as alternative introns in the 3′ untranslated region. These alternative splicing events target the resulting messenger RNAs for degradation by means of an RNA surveillance pathway called nonsense-mediated mRNA decay. Mouse orthologues of the human SR proteins exhibit the same unproductive splicing patterns. Three SR proteins have been previously shown to direct splicing of their own transcripts, and one of these is known to autoregulate its expression by coupling alternative splicing with decay[3–5]; our results suggest that unproductive splicing is important for regulation of the entire SR family. We find that unproductive splicing associated with conserved regions has arisen independently in different SR genes, suggesting that splicing factors may readily acquire this form of regulation.

The 11 human SR proteins comprise a homologous family of splicing factors (Supplementary Fig. 1) with diverse roles in RNA processing including control of export, translation, stability, and constitutive and alternative splicing[6]. To determine the extent of alternative splicing of SR genes themselves, we mined expressed sequence tag (EST) libraries for SR transcripts, aligned these ESTs to the human genome to infer splice junctions, and identified alternative junctions by comparison with a full-length reference mRNA of each gene. We found that all SR splicing regulators are alternatively spliced (Fig. 1, panel 2).

Notably, we found that all human SR genes had alternative splice forms that are expected to be degraded by nonsense-mediated mRNA decay (NMD) (Fig. 1, panel 2). The alternative splice forms contain stop codons that are thought to be recognized as premature by the NMD machinery because they are located more than 50 nucleotides upstream of the final exon–exon junction[7]. In some cases, ESTs with premature termination codons (PTCs) comprised a sizeable fraction of the total ESTs for that gene (Supplementary Table 3 and Supplementary Fig. 2). Seven of the SR genes exhibited a distinctive splicing pattern: an alternative, or cassette, exon was skipped in the reference isoform and included in an alternative isoform. *SRp20* (also called *SFRS3*), *SRP40* (*SFRS5*), *p54* (*SFRS11*), *SRP55* (*SFRS6*), *SRP75* (*SFRS4*) and *9G8* (*SFRS7*) contained poison cassette exons that introduced in-frame PTCs. *SRp38* (*FUSIP1*) had two linked alternative splicing events: an alternative 5′ splice site followed by a cassette exon, with a PTC in the extended upstream exon (Fig. 1, panel 2). For most of these genes, we also observed ESTs in which the introns flanking the poison cassette exons were retained and introduced PTCs. *SRp30c* (*SFRS9*) had similar intron-retention ESTs, but no ESTs supporting the cassette exon (discussed below). The remaining genes, *ASF/SF2* (also called *SFRS1*), *SC35* (*SFRS2*) and *SRP46* (*SFRS2B*), had alternative splicing in their 3′ untranslated regions (UTRs) (Fig. 1, panel 2). Although splicing in the 3′ UTR has no effect on the protein-coding sequence, the introduction of a new exon–exon junction more than 50 nucleotides downstream of the original stop codon marks that stop codon as premature and targets the transcript for NMD.

The existence of PTC^+ isoforms for all of the SR genes led us to ask whether this alternative splicing is truly unproductive, leading to downregulation of the spliced mRNA. We therefore measured the effect of NMD inhibition on mRNA isoform levels in HeLa cells using quantitative polymerase chain reaction with reverse transcription (quantitative RT–PCR). We disrupted NMD using short interfering (si)RNAs against *UPF1* (ref. 8), a key effector of the NMD pathway. If the PTCs trigger NMD as expected, the PTC^+ isoforms would be relatively stabilized upon UPF1 depletion. In all nine cases where the reference isoform could be detected, UPF1 depletion resulted in a 4- to 40-fold increase in the relative abundance of PTC-containing mRNAs (Fig. 2). We estimate that PTC^+ isoforms comprised about 2–14% of the spliced mRNA population from each gene in control conditions and 40–70% when NMD was inhibited (Supplementary Table 5), suggesting that a substantial fraction of SR transcripts is spliced into a form that is degraded by NMD. Time course experiments using cycloheximide rather than RNA interference to inhibit NMD showed similar effects on PTC^+ isoforms, supporting the expectation that the stabilization is a direct effect of NMD (Supplementary Fig. 4). Overall, our results indicate that alternative splice forms of the human SR protein family are produced at appreciable levels and are subsequently downregulated by NMD.

The 11 human SR genes correspond to 10 mouse SR genes (Supplementary Fig. 1; human *SRP46* is an expressed retropseudogene of *SC35*). We used ESTs to identify alternative splicing of the mouse SR genes. All of the genes except *SRp30c* had ESTs supporting the same unproductive splicing patterns in mouse as in human (Fig. 1, panel 3); the PTCs were in identical positions except in *SRP75* (called *Sfrs4* in mouse). Mouse *SRp30c* (also called *Sfrs9*) had a poison cassette exon, but there was no EST evidence of an equivalent cassette exon in

[1]Departments of Molecular and Cell Biology and [2]Plant and Microbial Biology, University of California, Berkeley, California 94720, USA. †Present address: Max Planck Institute for Evolutionary Anthropology, Leipzig D-04103, Germany.
*These authors contributed equally to this work.

the human orthologue (Fig. 1, panels 2 and 3). We identified a sequence within the human intron that was homologous to the mouse cassette exon and experimentally detected the predicted PTC$^+$ splice form in HeLa cells (Fig. 2). Because alternative splice forms are not generally shared between human and mouse, those that are conserved may be functionally important[9].

We next aligned the full genomic locus of each human SR protein with its mouse counterpart. Remarkably, all but one SR gene with conserved unproductive splice patterns had long regions of 100% nucleotide conservation between the human and mouse orthologues, ranging from 118 to 618 nucleotides in length (Fig. 1, panel 1). The conserved regions corresponded closely to the alternatively spliced poison cassette exons or to sequences flanking the 3′ UTR introns. The remaining gene, *SRp30c*, had shorter conserved stretches of 42 and 44 nucleotides at the ends of its poison cassette exon. The conserved sequences found in *9G8* (also called *Sfrs7* in mouse), *ASF/SF2* (*Sfrs1*), *SRp20* (*Sfrs3*), *p54* (*Sfrs11*) and *SRP55* (*Sfrs6*) were previously identified as ultraconserved genomic elements, defined as regions of at least 200 nucleotides of perfect identity between the human, mouse and rat genomes[1]; the conserved sequences in the other SR genes were shorter than that threshold, although still striking.

We compared the highly conserved regions among the human paralogues in order to investigate whether these regions arose from a common source. The most parsimonious scenario would be that the highly conserved sequences and their unproductive splicing

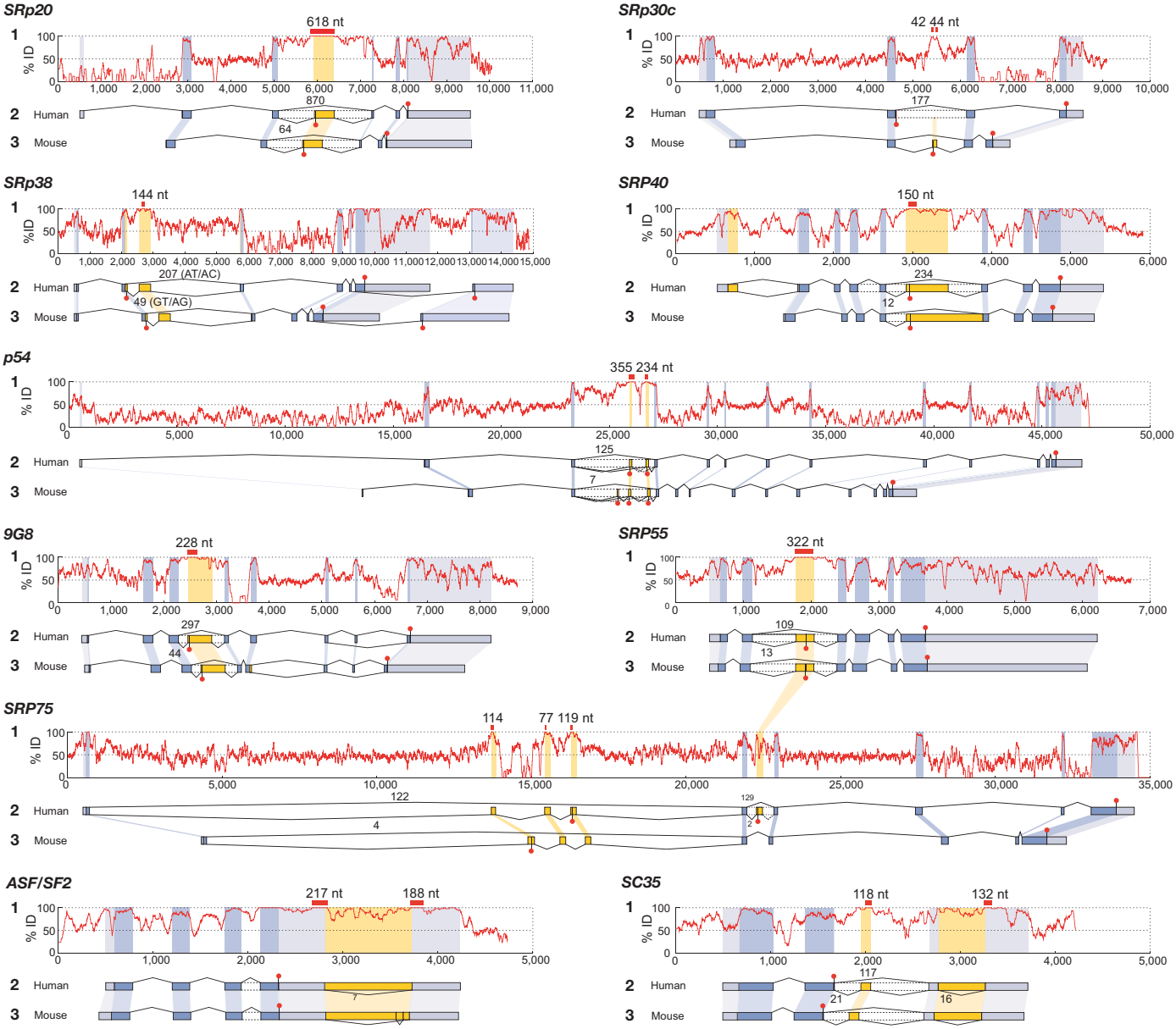


Figure 1 | SR gene alternative splice forms with PTCs are conserved between human and mouse and are associated with highly conserved and ultraconserved DNA elements. For each human/mouse SR gene pair, nucleotide conservation between human and mouse SR genomic loci is shown (panel 1). Per cent identity is shown in 50-nucleotide windows (*p54* in 100-nucleotide windows). DNA sequences with 100% conservation that are associated with alternatively spliced regions are marked (red bars). EST-inferred alternative splicing of human (panel 2) and mouse (panel 3) SR genes is shown with reference splice events (drawn above exon structure), alternative splicing (drawn below) and intron retention (dotted horizontal lines). Reference stop codons (above) and PTCs (below) are marked (red octagons). Gene regions are annotated as coding (blue), UTR (light blue) and alternatively spliced (orange). EST counts for alternative splice events and corresponding reference splice events are noted for human genes. The cassette exon in human *SRp30c* is shown by comparison to mouse. Sequence similarity between poison cassette exons of *SRP55* and *SRP75* is indicated with a connecting bar. nt, nucleotides.

originated in a common ancestor. Although some ultraconserved elements are known to have originated from a short interspersed element (SINE) retroposon[2], we found no evidence that the alternative regions of the SR genes contain known or novel repetitive elements (Supplementary Information). We also investigated whether a cassette exon was present in the ancestor of the SR paralogues. We found that the poison cassette exons were in non-homologous positions in different SR genes (Fig. 1, panel 2), implying independent rather than common origins. Moreover, local sequence alignments showed no significant nucleotide identity between the alternative regions of most human SR genes (Supplementary Information). The only significant relationship among alternative regions was found between the poison cassette exons in reference intron 2 of the most closely related human SR genes, *SRP55* and *SRP75*; this exon may therefore have been present in their common ancestor (Fig. 1, panel 2). This exon was ultraconserved between human and mouse in *SRP55* but not in *SRP75*, and it was observed rarely in *SRP75* ESTs, indicating that the exon may be under less selective pressure in *SRP75*. *SRP75* had additional, highly conserved poison cassette exons in its first reference intron that were not found in *SRP55*. Further study of the evolutionary dynamics of these ultraconserved regions may offer an insight into their origin and function. It is remarkable that similar splicing associated with highly conserved sequences seems to have arisen independently in different SR genes within this family.

Commonly measured constraints on sequence evolution do not explain the extreme conservation in these genes. Although a fraction

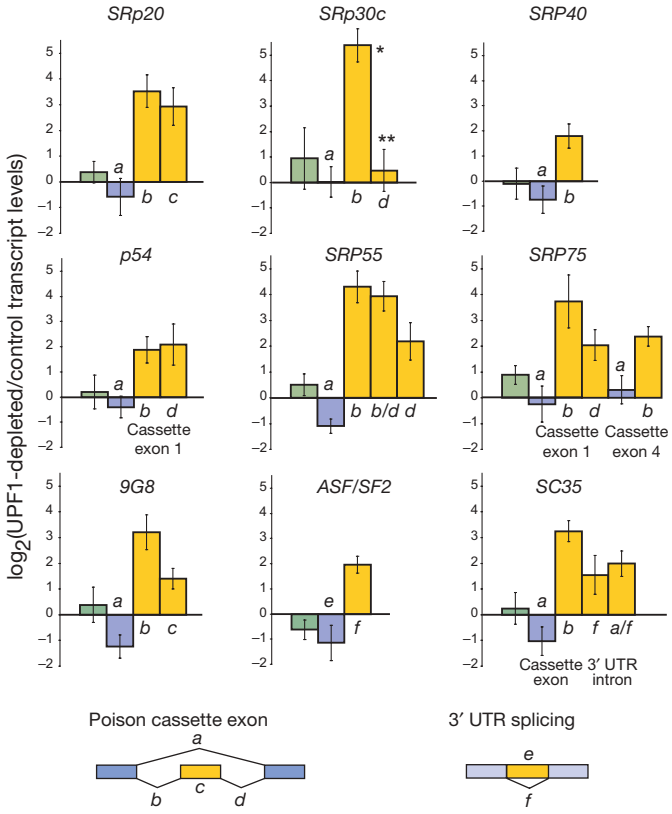


Figure 2 | SR gene alternative isoforms containing PTCs are degraded by NMD. Quantitative RT–PCR was used to assay levels of mRNAs in HeLa cells treated with control or *UPF1* siRNAs. The ratio of NMD-inhibited to control transcript levels was measured for constitutive (green), reference (blue; *a*, *e*) and PTC$^+$ (orange; *b*–*d*, *f*) regions. The data are $\log_2$ means of transcript ratios (±s.d.; three PCR replicates of four independent transfections). Asterisk, *SRp30c* PTC$^+$ isoform levels in control cells were at the lower limit of detection. Double asterisk, junction probes for *SRp30c* are predicted by homology to mouse ESTs and may not detect the actual junction used in human.

of the PTC$^+$ isoforms might escape NMD and be translated into truncated protein, selective pressure on the conserved regions is unlikely to arise from protein-coding constraints[3,10]. The stop codons are generally near the beginning of the conserved regions, or entirely upstream of the conserved region in the case of UTR splicing, leaving up to 422 nucleotides of non-coding, 100% conserved sequence downstream of the stop codon. A more probable explanation for the conservation is the presence of regulatory sequences that might control the use of the alternative splice sites. We found no consistent over-representation of binding sites for four well-characterized SR proteins in or near the alternative regions, and known exonic splicing enhancers were under-represented in the cassette exons (Supplementary Information); nonetheless, other binding sites may be present[11,12]. Similarly, *trans*-acting RNAs such as microRNAs might bind to the alternative regions. We identified putative microRNA-binding sites in the SR mRNAs[30], but the sites are not over-represented within the conserved alternative regions (Supplementary Information). The highly conserved sequences might also contribute to regulation through RNA structure, but most of the ultraconserved elements previously examined lack significantly low-energy secondary structure[1]. Extending the conservation analysis to other vertebrate genomes may help to pinpoint the functionally essential sequence elements[13]. We suspect that intertwined and as yet unknown constraints from RNA structure, RNA binding and protein binding may be at play.

The frequency of alternative splicing, the similar splice patterns and the exceptional sequence conservation of the SR genes strongly suggest that their unproductive splicing is functionally important. The human SR protein SC35 provides one model for this function. *SC35* pre-mRNAs can be spliced into either a reference isoform or PTC$^+$ isoforms. When SC35 protein is overproduced, it affects the splicing of its own pre-mRNAs, increasing the proportion of NMD-target isoforms[5]. This creates a negative feedback loop: high levels of SC35 protein cause a greater fraction of *SC35* transcripts to be spliced into the unproductive isoform and then degraded, thereby down-regulating SC35 protein production. Our data show that the entire SR family might couple alternative splicing and NMD to regulate protein production, a mechanism termed regulated unproductive splicing and translation (RUST)[14]. Consistent with this model, two other human SR proteins, 9G8 and SRp20, are known to affect splicing of their own transcripts[3,4]. Our data also lend support to this model: after NMD inhibition, some reference splice forms decrease in abundance (Fig. 2). Because inhibiting NMD should not directly affect the abundance of productive mRNAs, this decrease indicates a change in the underlying splicing ratio of productive to unproductive isoforms, consistent with negative feedback regulation. Other modes of regulation, such as cross-regulation, may also be involved. Notably, SR proteins are also regulated by Clk kinases, which themselves have PTC$^+$ splice forms. Modulation of SR proteins with a network of unproductive splicing might affect a wide range of downstream splicing targets[15–17].

The association between unproductive splicing and ultraconserved regions extends beyond the SR family to other RNA-binding proteins. Two SR-related genes and six heterogeneous nuclear ribonucleoprotein (hnRNP) genes also contain ultraconserved elements, and one hnRNP, PTB (also called PTBP1), autoregulates its expression by means of tissue-regulated unproductive splicing[1,18,19]. Furthermore, this mode of regulation may be phylogenetically widespread, as SR genes are alternatively spliced in other species including nematodes, flies and plants[20–24]. Some of these splice forms are homologous to human splice forms, such as a poison cassette exon found in the *Ciona intestinalis* orthologue of *SRp20* (Supplementary Fig. 5). Although the poison cassette splicing is conserved, the sequence ultraconservation does not extend from human to *Ciona*, showing that the regulatory phenomenon may be conserved but, if so, the specific sequence constraints are flexible.

What benefit would regulation by unproductive splicing confer, and why did this particular mechanism evolve in SR genes? The cell

may use unproductive splicing to maintain homeostasis of splicing factors. An SR-related splicing factor has been shown to have unusually low cell-to-cell variability in protein level[25]. Unproductive splicing may also allow tissue-specific expression of SR proteins[4]. If regulation would provide a selective advantage, alternative splicing linked with NMD is an evolutionarily accessible means of regulation for splicing factors, as they are inherently capable of binding mRNA. Whereas unproductive splice forms of many genes may represent cellular noise[26], the splicing events we have identified demonstrate how unproductive splicing can provide the capacity for regulation. Our observations paint a picture of unproductive splicing associated with highly conserved sequences as a regulatory process that has evolved repeatedly, affecting expression throughout an entire family of splicing factors.

METHODS

Detailed methods are supplied in Supplementary Information. To identify alternative splice forms of the SR genes, we aligned ESTs from UniGene and reference mRNAs to genomic loci using spidey v1.4, and compared the EST-inferred and reference splice junctions[27]. Alternative splice events seen in only one EST or occurring within ten nucleotides of a more prevalent splice event were discarded. To test whether alternative splice forms are degraded by NMD, we transfected HeLa cells with plasmids expressing siRNAs against *UPF1* (ref. 8) and reduced the level of UPF1 protein to $<5\%$ of the control level (Supplementary Fig. 3). For each SR gene, we measured relative mRNA levels using quantitative RT–PCR with primers to detect splice junctions or exons found in reference, PTC^+ and constitutive splice forms. The fold change in levels between control and experimental conditions for each primer pair was normalized to the average of three control genes: β-actin, *SDHA* and *TBP*.

Genomic alignments were constructed by splitting the human and mouse SR gene loci into individual exon and intron sequences and aligning each pair of orthologous sequences using the global alignment program from FASTA v2.1 (ref. 28). We searched for known and novel repeat elements within the genes using the RepeatMasker and chained self-assembly tracks of the UCSC human genome browser[29]. We performed local alignments to compare the poison cassette exons plus their flanking introns between the different human SR paralogues. The ESEfinder web server was used to search each SR gene containing a poison cassette exon for sites scoring higher than the default thresholds of the ASF/SF2, SC35, SRP40 and SRP55 binding motifs, and the RESCUE-ESE web server was used to search for known exonic splicing enhancer (ESE) sequences[11,12]. miRanda v1.0b was used to identify potential microRNA target sites[30].

Received 16 October 2006; accepted 28 February 2007.
Published online 14 March 2007.

1. Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304,** 1321–1325 (2004).
2. Bejerano, G. *et al.* A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441,** 87–90 (2006).
3. Jumaa, H. & Nielsen, P. J. The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.* **16,** 5077–5085 (1997).
4. Lejeune, F., Cavaloc, Y. & Stevenin, J. Alternative splicing of intron 3 of the serine/arginine-rich protein 9G8 gene. Identification of flanking exonic splicing enhancers and involvement of 9G8 as a trans-acting factor. *J. Biol. Chem.* **276,** 7850–7858 (2001).
5. Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J. & Soret, J. SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J.* **20,** 1785–1796 (2001).
6. Sanford, J. R., Ellis, J. & Caceres, J. F. Multiple roles of arginine/serine-rich splicing factors in RNA processing. *Biochem. Soc. Trans.* **33,** 443–446 (2005).
7. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* **23,** 198–199 (1998).
8. Paillusson, A., Hirschi, N., Vallan, C., Azzalin, C. M. & Muhlemann, O. A GFP-based reporter system to monitor nonsense-mediated mRNA decay. *Nucleic Acids Res.* **33,** e54 (2005).
9. Lareau, L. F., Green, R. E., Bhatnagar, R. S. & Brenner, S. E. The evolving roles of alternative splicing. *Curr. Opin. Struct. Biol.* **14,** 273–282 (2004).
10. Ge, H., Zuo, P. & Manley, J. L. Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell* **66,** 373–382 (1991).
11. Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31,** 3568–3571 (2003).
12. Fairbrother, W. G., Yeh, R. F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297,** 1007–1013 (2002).
13. Hughes, J. R. *et al.* Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *Proc. Natl Acad. Sci. USA* **102,** 9830–9835 (2005).
14. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100,** 189–192 (2003).
15. Hillman, R. T., Green, R. E. & Brenner, S. E. An unappreciated role for RNA surveillance. *Genome Biol.* **5,** R8 (2004).
16. Duncan, P. I., Stojdl, D. F., Marius, R. M. & Bell, J. C. *In vivo* regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. *Mol. Cell. Biol.* **17,** 5996–6001 (1997).
17. Garcia-Sacristan, A., Fernandez-Nestosa, M. J., Hernandez, P., Schvartzman, J. B. & Krimer, D. B. Protein kinase clk/STY is differentially regulated during erythroleukemia cell differentiation: a bias toward the skipped splice variant characterizes postcommitment stages. *Cell Res.* **15,** 495–503 (2005).
18. Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A. & Smith, C. W. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* **13,** 91–100 (2004).
19. Yeo, G. W., Van Nostrand, E., Holste, D., Poggio, T. & Burge, C. B. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc. Natl Acad. Sci. USA* **102,** 2850–2855 (2005).
20. Kumar, S. & Lopez, A. J. Negative feedback regulation among SR splicing factors encoded by Rbp1 and Rbp1-like in *Drosophila*. *EMBO J.* **24,** 2646–2655 (2005).
21. Zorio, D. A., Lea, K. & Blumenthal, T. Cloning of *Caenorhabditis* $U2AF^{65}$: an alternatively spliced RNA containing a novel exon. *Mol. Cell. Biol.* **17,** 946–953 (1997).
22. Morrison, M., Harris, K. S. & Roth, M. B. *smg* mutants affect the expression of alternatively spliced SR protein mRNAs in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **94,** 9782–9785 (1997).
23. Kalyna, M., Lopato, S., Voronin, V. & Barta, A. Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.* **34,** 4395–4405 (2006).
24. Iida, K. & Go, M. Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. *Mol. Biol. Evol.* **23,** 1085–1094 (2006).
25. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444,** 643–646 (2006).
26. Pan, Q. *et al.* Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev.* **20,** 153–158 (2006).
27. Wheeler, D. L. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **34,** D173–D180 (2006).
28. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85,** 2444–2448 (1988).
29. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).
30. John, B. *et al.* Human microRNA targets. *PLoS Biol.* **2,** e363 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank present and past members of the Brenner laboratory and J. Pleiss for discussions and comments on the manuscript, Q. Meng and D. Rio for expertise, A. Fisher for assistance, and the Feldman, Luan, Rio and Tjian laboratories for use of their equipment. This work was supported by NIH grants and a Sloan fellowship to S.E.B. L.F.L. was supported by an NIH training grant. R.E.G. is supported by an NSF postdoctoral fellowship in Biological Informatics.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.E.B. (brenner@compbio.berkeley.edu).

ARTICLES

An experimental test of non-local realism

Simon Gröblacher[1,2], Tomasz Paterek[3,4], Rainer Kaltenbaek[1], Časlav Brukner[1,2], Marek Żukowski[1,3], Markus Aspelmeyer[1,2] & Anton Zeilinger[1,2]

Most working scientists hold fast to the concept of 'realism'—a viewpoint according to which an external reality exists independent of observation. But quantum physics has shattered some of our cornerstone beliefs. According to Bell's theorem, any theory that is based on the joint assumption of realism and locality (meaning that local events cannot be affected by actions in space-like separated regions) is at variance with certain quantum predictions. Experiments with entangled pairs of particles have amply confirmed these quantum predictions, thus rendering local realistic theories untenable. Maintaining realism as a fundamental concept would therefore necessitate the introduction of 'spooky' actions that defy locality. Here we show by both theory and experiment that a broad and rather reasonable class of such non-local realistic theories is incompatible with experimentally observable quantum correlations. In the experiment, we measure previously untested correlations between two entangled photons, and show that these correlations violate an inequality proposed by Leggett for non-local realistic theories. Our result suggests that giving up the concept of locality is not sufficient to be consistent with quantum experiments, unless certain intuitive features of realism are abandoned.

Physical realism suggests that the results of observations are a consequence of properties carried by physical systems. It remains surprising that this tenet is very little challenged, as its significance goes far beyond science. Quantum physics, however, questions this concept in a very deep way. To maintain a realistic description of nature, non-local hidden-variable theories are being discussed as a possible completion of quantum theory. They offer to explain intrinsic quantum phenomena—above all, quantum entanglement[1]—by non-local influences. Up to now, however, it has not been possible to test such theories in experiments. We present an inequality, similar in spirit to the seminal one given by Clauser, Horne, Shimony and Holt[2] on local hidden variables, that allows us to test an important class of non-local hidden-variable theories against quantum theory. The theories under test provide an explanation of all existing two-qubit Bell-type experiments. Our derivation is based on a recent incompatibility theorem by Leggett[3], which we extend so as to make it applicable to real experimental situations and also to allow simultaneous tests of all local hidden-variable models. Finally, we perform an experiment that violates the new inequality and hence excludes for the first time a broad class of non-local hidden-variable theories.

Quantum theory gives only probabilistic predictions for individual events. Can one go beyond this? Einstein's view[4,5] was that quantum theory does not provide a complete description of physical reality: ''While we have thus shown that the wavefunction does not provide a complete description of the physical reality, we left open the question of whether or not such a description exists. We believe, however, that such a theory is possible.''[4]. It remained an open question whether the theory could be completed in Einstein's sense[6]. If so, more complete theories based on objective properties of physical systems should be possible. Such models are referred to as hidden-variable theories.

Bell's theorem[7] proves that all hidden-variable theories based on the joint assumption of locality and realism are at variance with the predictions of quantum physics. Locality prohibits any influences between events in space-like separated regions, while realism claims that all measurement outcomes depend on pre-existing properties of objects that are independent of the measurement. The more refined versions of Bell's theorem by Clauser, Horne, Shimony and Holt[2] and by Clauser and Horne[8,9] start from the assumptions of local realism and result in inequalities for a set of statistical correlations (expectation values), which must be satisfied by all local realistic hidden-variable theories. The inequalities are violated by quantum mechanical predictions. Greenberger, Horne and Zeilinger[10,11] showed that already perfect correlations of systems with at least three particles are inconsistent with these assumptions. So far, all experiments motivated by these theorems are in full agreement with quantum predictions[12–17]. For some time, loopholes existed that allowed the observed correlations to be explained within local realistic theories. In particular, an ideal Bell experiment has to be performed with detectors of sufficiently high efficiency (to close the 'detection loophole') and with experimental settings that are randomly chosen in space-like separated regions (to close the 'locality loophole'). Since the first successful Bell experiment by Freedman and Clauser[12], later implementations have continuously converged to closing both the locality loophole[14,15,18,19] on the one hand and the detection loophole[16,20] on the other hand. Therefore it is reasonable to consider the violation of local realism a well established fact.

The logical conclusion one can draw from the violation of local realism is that at least one of its assumptions fails. Specifically, either locality or realism or both cannot provide a foundational basis for quantum theory. Each of the resulting possible positions has strong supporters and opponents in the scientific community. However, Bell's theorem is unbiased with respect to these views: on the basis of this theorem, one cannot, even in principle, favour one over the other. It is therefore important to ask whether incompatibility theorems similar to Bell's can be found in which at least one of these concepts is relaxed. Our work addresses a broad class of non-local hidden-variable theories that are based on a very plausible type of realism and that provide an explanation for all existing Bell-type experiments. Nevertheless we demonstrate, both in theory and experiment, their conflict with quantum predictions and observed

[1]Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria. [2]Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmanngasse 3, A-1090 Vienna, Austria. [3]Institute of Theoretical Physics and Astrophysics, University of Gdansk, ul. Wita Stwosza 57, PL-08-952 Gdansk, Poland. [4]The Erwin Schrödinger International Institute for Mathematical Physics (ESI), Boltzmanngasse 9, A-1090 Vienna, Austria.

measurement data. Following the recent approach of Leggett[3], who introduced the class of non-local models and formulated an incompatibility theorem, we have analysed its assumptions and derived an inequality valid for such theories that can be experimentally tested. In addition, the experiments allow for a simultaneous test of all local hidden-variable models—that is, the measurement data can neither be explained by a local realistic model nor by the considered class of non-local models.

The theories under investigation describe experiments on pairs of particles. It is sufficient for our purposes to discuss two-dimensional quantum systems. We will hence focus our description on the polarization degree of freedom of photons. The theories are based on the following assumptions: (1) all measurement outcomes are determined by pre-existing properties of particles independent of the measurement (realism); (2) physical states are statistical mixtures of subensembles with definite polarization, where (3) polarization is defined such that expectation values taken for each subensemble obey Malus' law (that is, the well-known cosine dependence of the intensity of a polarized beam after an ideal polarizer).

These assumptions are in a way appealing, because they provide a natural explanation of quantum mechanically separable states (polarization states indeed obey Malus' law). In addition, they do not explicitly demand locality; that is, measurement outcomes may very well depend on parameters in space-like separated regions. As a consequence, such theories can explain important features of quantum mechanically entangled (non-separable) states of two particles (a specific model can be found in Supplementary Information): first, they do not allow information to be transmitted faster than the speed of light; second, they reproduce perfect correlations for all measurements in the same bases, which is a fundamental feature of the Bell singlet state; and third, they provide a model for all thus far performed experiments in which the Clauser, Horne, Shimony and Holt (CHSH) inequality was violated. Nevertheless, we will show that all models based on assumptions (1)–(3) are at variance with other quantum predictions.

A general framework of such models is the following: assumption (1) requires that an individual binary measurement outcome A for a polarization measurement along direction $\mathbf{a}$ (that is, whether a single photon is transmitted or absorbed by a polarizer set at a specific angle) is predetermined by some set of hidden variables λ, and a three-dimensional vector $\mathbf{u}$, as well as by some set of other possibly non-local parameters η (for example, measurement settings in space-like separated regions)—that is, $A = A(\lambda,\mathbf{u},\mathbf{a},\eta)$. According to assumption (3), particles with the same $\mathbf{u}$ but with different λ build up subensembles of 'definite polarization' described by a probability distribution $\rho_{\mathbf{u}}(\lambda)$. The expectation value $\bar{A}(\mathbf{u})$, obtained by averaging over λ, fulfils Malus' law, that is, $\bar{A}(\mathbf{u}) = \int \mathrm{d}\lambda \rho_{\mathbf{u}}(\lambda) A(\lambda,\mathbf{u},\mathbf{a},\eta) = \mathbf{u}\cdot\mathbf{a}$. Finally, with assumption (2), the measured expectation value for a general physical state is given by averaging over the distribution $F(\mathbf{u})$ of subensembles, that is, $\langle A\rangle = \int \mathrm{d}\mathbf{u} F(\mathbf{u})\overline{A}(\mathbf{u})$.

Let us consider a specific source, which emits pairs of photons with well-defined polarizations $\mathbf{u}$ and $\mathbf{v}$ to laboratories of Alice and Bob, respectively. The local polarization measurement outcomes A and B are fully determined by the polarization vector, by an additional set of hidden variables λ specific to the source and by any set of parameters η outside the source. For reasons of clarity, we choose an explicit non-local dependence of the outcomes on the settings $\mathbf{a}$ and $\mathbf{b}$ of the measurement devices. Note, however, that this is just an example of a possible non-local dependence, and that one can choose any other set out of η. Each emitted pair is fully defined by the subensemble distribution $\rho_{\mathbf{u},\mathbf{v}}(\lambda)$. In agreement with assumption (3) we impose the following conditions on the predictions for local averages of such measurements (all polarizations and measurement directions are represented as vectors on the Poincaré sphere[21]):

$$\bar{A}(\mathbf{u}) = \int \mathrm{d}\lambda \rho_{\mathbf{u},\mathbf{v}}(\lambda) A(\mathbf{a},\mathbf{b},\lambda) = \mathbf{u}\cdot\mathbf{a} \quad (1)$$

$$\overline{B}(\mathbf{v}) = \int \mathrm{d}\lambda \rho_{\mathbf{u},\mathbf{v}}(\lambda) B(\mathbf{b},\mathbf{a},\lambda) = \mathbf{v}\cdot\mathbf{b} \quad (2)$$

It is important to note that the validity of Malus' law imposes the non-signalling condition on the investigated non-local models, as the local expectation values do only depend on local parameters. The correlation function of measurement results for a source emitting well-polarized photons is defined as the average of the products of the individual measurement outcomes:

$$\overline{AB}(\mathbf{u},\mathbf{v}) = \int \mathrm{d}\lambda \rho_{\mathbf{u},\mathbf{v}}(\lambda) A(\mathbf{a},\mathbf{b},\lambda) B(\mathbf{b},\mathbf{a},\lambda) \quad (3)$$

For a general source producing mixtures of polarized photons the observable correlations are averaged over a distribution of the polarizations $F(\mathbf{u},\mathbf{v})$, and the general correlation function E is given by:

$$E = \langle AB\rangle = \int \mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} F(\mathbf{u},\mathbf{v})\overline{AB}(\mathbf{u},\mathbf{v}) \quad (4)$$

It is a very important trait of this model that there exist subensembles of definite polarizations (independent of measurements) and that the predictions for the subensembles agree with Malus' law. It is clear that other classes of non-local theories, possibly even fully compliant with all quantum mechanical predictions, might exist that do not have this property when reproducing entangled states. Such theories may, for example, include additional communication[22] or dimensions[23]. A specific case deserving comment is Bohm's theory[24]. There the non-local correlations are a consequence of the non-local quantum potential, which exerts suitable torque on the particles leading to experimental results compliant with quantum mechanics. In that theory, neither of the two particles in a maximally entangled state carries any angular momentum at all when emerging from the source[25]. In contrast, in the Leggett model, it is the total ensemble emitted by the source that carries no angular momentum, which is a consequence of averaging over the individual particles' well defined angular momenta (polarization).

The theories described here are incompatible with quantum theory. The basic idea of the incompatibility theorem[3] uses the following identity, which holds for any numbers $A = \pm 1$ and $B = \pm 1$:

$$-1 + |A+B| = AB = 1 - |A-B| \quad (5)$$

One can apply this identity to the dichotomic measurement results $A = A(\mathbf{a},\mathbf{b},\lambda) = \pm 1$ and $B = B(\mathbf{b},\mathbf{a},\lambda) = \pm 1$. The identity holds even if the values of A and B mutually depend on each other. For example, the value of a specific outcome A can depend on the value of an actually obtained result B. In contrast, in the derivation of the CHSH inequality it is necessary to assume that A and B do not depend on each other. Therefore, any kind of non-local dependencies used in the present class of theories are allowed. Taking the average over the subensembles with definite polarizations we obtain:

$$-1 + \int \mathrm{d}\lambda \rho_{\mathbf{u},\mathbf{v}}(\lambda)|A+B| = \int \mathrm{d}\lambda \rho_{\mathbf{u},\mathbf{v}}(\lambda) AB = 1 - \int \mathrm{d}\lambda \rho_{\mathbf{u},\mathbf{v}}(\lambda)|A-B| \quad (6)$$

Denoting these averages by bars, one arrives at the shorter expression:

$$-1 + \overline{|A+B|} = \overline{AB} = 1 - \overline{|A-B|} \quad (7)$$

As the average of the modulus is greater than or equal to the modulus of the averages, one gets the set of inequalities:

$$-1 + \left|\overline{A}+\overline{B}\right| \le \overline{AB} \le 1 - \left|\overline{A}-\overline{B}\right| \quad (8)$$

By inserting Malus' law, equations (1) and (2), in equation (8), and by using expression (4), one arrives at a set of inequalities for experimentally accessible correlation functions (for a detailed derivation, see the Supplementary Information). In particular, if we let Alice

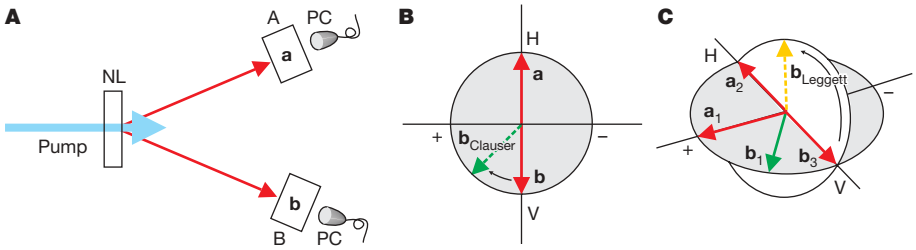


Figure 1 | Testing non-local hidden-variable theories. A, Diagram of a standard two-photon experiment to test for hidden-variable theories. When pumping a nonlinear crystal (NL) with a strong pump field, photon pairs are created via SPDC and their polarization is detected with single-photon counters (PC). Local measurements at A and B are performed along directions **a** and **b** on the Poincaré sphere, respectively. Depending on the measurement directions, the obtained correlations can be used to test Bell inequalities (**B**) or Leggett-type inequalities (**C**). **B**, Correlations in one plane. Shown are measurements along directions in the linear plane of the Poincaré sphere (H (V) denotes horizontal (vertical) polarization). The original experiments by Wu and Shaknov[26] and Kocher and Commins[27], designed to test quantum predictions for correlated photon pairs, measured perfect correlations (solid lines). Measurements along the dashed line allow a Bell test, as was first performed by Freedman and Clauser[12]. **C**, Correlations in orthogonal planes. All current experimental tests to violate Bell's inequality (CHSH) are performed within the shaded plane. Out-of plane measurements are required for a direct test of the class of non-local hidden-variable theories, as was first suggested by Leggett.

choose her observable from the set of two settings $\mathbf{a}_1$ and $\mathbf{a}_2$, and Bob from the set of three settings $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3 = \mathbf{a}_2$, the following generalized Leggett-type inequality is obtained:

$$S_{\mathrm{NLHV}} = |E_{11}(\varphi) + E_{23}(0)| + |E_{22}(\varphi) - E_{23}(0)| \leq 4 - \frac{4}{\pi}\left|\sin\frac{\varphi}{2}\right| \quad (9)$$

where $E_{kl}(\varphi)$ is a uniform average of all correlation functions, defined in the plane of $\mathbf{a}_k$ and $\mathbf{b}_l$ with the same relative angle φ; the subscript NLHV stands for 'non-local hidden variables'. For the inequality to be applied, vectors $\mathbf{a}_1$ and $\mathbf{b}_1$ necessarily have to lie in a plane orthogonal to the one defined by $\mathbf{a}_2$ and $\mathbf{b}_2$. This contrasts with the standard experimental configuration used to test the CHSH inequality, which is maximally violated for settings in one plane.

The situation resembles in a way the status of the Einstein, Podolsky and Rosen (EPR) paradox before the advent of Bell's theorem and its first experimental tests. The experiments of Wu and Shaknov[26] and of Kocher and Commins[27] were designed to demonstrate the validity of a quantum description of photon-pair correlations. As this task only required the testing of correlations along the same polarization direction, their results could not provide experimental data for the newly derived Bell inequalities (Fig. 1A, B). Curiously, as was shown by Clauser, Horne, Shimony and Holt, only a small modification of the measurement directions, such that non-perfect correlations of an entangled state are probed, was sufficient to test Bell's inequalities. The seminal experiment by Freedman and Clauser[12] was the first direct and successful test[28]. Today, all Bell tests—that is, tests of local realism—are performed by testing correlations of measurements along directions that lie in the same plane of the Poincaré sphere. Similar to the previous case, violation of the Leggett-type inequality requires only small modifications to that arrangement: To test the inequality, correlations of measurements along two orthogonal planes have to be probed (Fig. 1C). Therefore the existing data of all Bell tests cannot be used to test the class of non-local theories considered here.

Quantum theory violates inequality (9). Consider the quantum predictions for the polarization singlet state of two photons,

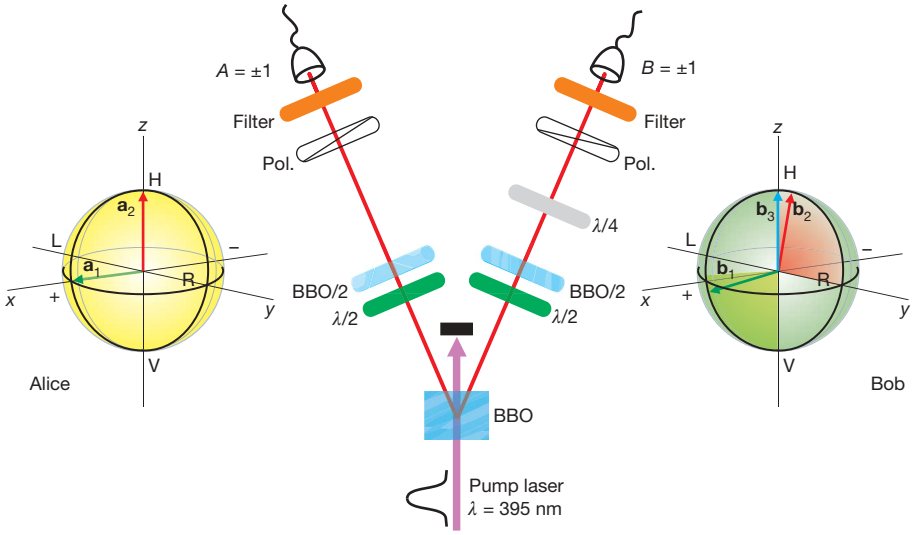


Figure 2 | Experimental set-up. A 2-mm-thick type-II β-barium-borate (BBO) crystal is pumped with a pulsed frequency-doubled Ti:sapphire laser (180 fs) at $\lambda = 395$ nm wavelength and ~150 mW optical c.w. power. The crystal is aligned to produce the polarization-entangled singlet state

$$|\Psi^-\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{2}}\left[|H\rangle_{\mathrm{A}}|V\rangle_{\mathrm{B}} - |V\rangle_{\mathrm{A}}|H\rangle_{\mathrm{B}}\right].$$

Spatial and temporal distinguishability of the produced photons (induced by birefringence in the BBO) are compensated by a combination of half-wave plates ($\lambda/2$) and additional BBO crystals (BBO/2), while spectral distinguishability (due to the broad spectrum of the pulsed pump) is eliminated by narrow spectral filtering of 1 nm bandwidth in front of each detector. In addition, the reduced pump power diminishes higher-order SPDC emissions of multiple photon pairs. This allows us to achieve a two-photon visibility of about 99%, which is well beyond the required threshold of 97.4%. The arrows in the Poincaré spheres indicate the measurement settings of Alice's and Bob's polarizers for the maximal violation of inequality (9). Note that setting $\mathbf{b}_2$ lies in the y–z plane and therefore a quarter-wave plate has to be introduced on Bob's side. The coloured planes indicate the measurement directions for various difference angles φ for both inequalities.

$|\Psi^-\rangle_{AB} = \frac{1}{\sqrt{2}}\left[|H\rangle_A|V\rangle_B - |V\rangle_A|H\rangle_B\right]$, where, for example, $|H\rangle_A$ denotes a horizontally polarized photon propagating to Alice. The quantum correlation function for the measurements $\mathbf{a}_k$ and $\mathbf{b}_l$ performed on photons depends only on the relative angles between these vectors, and therefore $E_{kl} = -\mathbf{a}_k \cdot \mathbf{b}_l = -\cos\varphi$. Thus the left hand side of inequality (9), for quantum predictions, reads $|2(\cos\varphi + 1)|$. The maximal violation of inequality (9) is for $\varphi_{max} = 18.8^\circ$. For this difference angle, the bound given by inequality (9) equals 3.792 and the quantum value is 3.893.

Although this excludes the non-local models, it might still be possible that the obtained correlations could be explained by a local realistic model. In order to avoid that, we have to exclude both local realistic and non-local realistic hidden-variable theories. Note however that such local realistic theories need not be constrained by assumptions (1)–(3). The violation of the CHSH inequality invalidates all local realistic models. If one takes

$$S_{CHSH} = |E_{11} + E_{12} - E_{21} + E_{22}| \leq 2 \qquad (10)$$

the quantum value of the left hand side for the settings used to maximally violate inequality (9) is 2.2156.

The correlation function determined in an actual experiment is typically reduced by a visibility factor V to $E^{exp} = -V\cos\varphi$ owing to noise and imperfections. Thus to observe violations of inequality (9) (and inequality (10)) in the experiment, one must have a sufficiently high experimental visibility of the observed interference. For the optimal difference angle $\varphi_{max} = 18.8^\circ$, the minimum required visibility is given by the ratio of the bound (3.792) and the quantum value (3.893) of inequality (9), or ~97.4%. We note that in standard Bell-type experiments, a minimum visibility of only ~71% is sufficient to violate the CHSH inequality, inequality (10), at the optimal settings. For the settings used here, the critical visibility reads $2/2.2156 \approx 90.3\%$, which is much lower than 97.4%.

In the experiment (see Fig. 2), we generate pairs of polarization entangled photons via spontaneous parametric down-conversion (SPDC). The photon source is aligned to produce pairs in the polarization singlet state. We observed maximal coincidence count rates (per 10 s), in the H/V basis, of around 3,500 with single count rates of 95,000 (Alice) and 105,000 (Bob), 3,300 coincidences in the $\pm 45^\circ$ basis (75,000 singles at Alice and 90,000 at Bob), and 2,400 coincidences in the R/L basis (70,000 singles at Alice and 70,000 at Bob). The reduced count rates in the R/L basis are due to additional retarding elements in the beam path. The two-photon visibilities are approximately $99.0 \pm 1.2\%$ in the H/V basis, $99.2 \pm 1.6\%$ in the $\pm 45^\circ$ basis and $98.9 \pm 1.7\%$ in the R/L basis, which—to our knowledge—is the highest reported visibility for a pulsed SPDC scheme. So far, no experimental evidence against the rotational invariance of the singlet state exists. We therefore replace the rotation averaged correlation functions in inequality (9) with their values measured for one pair of settings (in the given plane).

In terms of experimental count rates, the correlation function $E(\mathbf{a},\mathbf{b})$ for a given pair of general measurement settings is defined by

$$E(\mathbf{a},\mathbf{b}) = \frac{N_{++} + N_{--} - N_{+-} - N_{-+}}{N_{++} + N_{--} + N_{+-} + N_{-+}} \qquad (11)$$

where N_{AB} denotes the number of coincident detection events between Alice's and Bob's measurements within the integration time. We ascribe the number +1, if Alice (Bob) detects a photon polarized along $\mathbf{a}$ ($\mathbf{b}$), and −1 for the orthogonal direction $\mathbf{a}^\perp$ ($\mathbf{b}^\perp$). For example, N_{+-} denotes the number of coincidences in which Alice obtains $\mathbf{a}$ and Bob $\mathbf{b}^\perp$. Note that $E(\mathbf{a}_k,\mathbf{b}_l) = E_{kl}(\varphi)$, where φ is the difference angle between the vectors $\mathbf{a}$ and $\mathbf{b}$ on the Poincaré sphere.

To test inequality (9), three correlation functions ($E(\mathbf{a}_1,\mathbf{b}_1)$, $E(\mathbf{a}_2,\mathbf{b}_2)$, $E(\mathbf{a}_2,\mathbf{b}_3)$) have to be extracted from the measured data. We choose observables $\mathbf{a}_1$ and $\mathbf{b}_1$ as linear polarization measurements (in the x–z plane on the Poincaré sphere; see Fig. 2) and $\mathbf{a}_2$ and $\mathbf{b}_2$ as elliptical polarization measurements in the y–z plane. Two further correlation functions ($E(\mathbf{a}_2,\mathbf{b}_1)$ and $E(\mathbf{a}_1,\mathbf{b}_2)$) are extracted to test the CHSH inequality, inequality (10).

The first set of correlations, in the x–z plane, is obtained by using linear polarizers set to α_1 and β_1 (relative to the z axis) at Alice's and Bob's location, respectively. In particular, $\alpha_1 = \pm 45^\circ$, while β_1 is chosen to lie between 45° and 160° (green arrows in Fig. 2). The second set of correlations (necessary for CHSH) is obtained in the same plane for $\alpha_2 = 0^\circ/90^\circ$ and β_1 between 45° and 160°. The set of correlations for measurements in the y–z plane is obtained by introducing a quarter-wave plate with the fast axis aligned along the (horizontal) 0°-direction at Bob's site, which effectively rotates the polarization state by 90° around the z-axis on the Poincaré sphere (red arrows in Fig. 2). The polarizer angles are then set to $\alpha_2 = 0^\circ/90^\circ$ and β_2 is scanned between 0° and 115°. With the same β_2 and $\alpha_1 = \pm 45^\circ$, the expectation values specific only for the CHSH case are measured. The remaining measurement for inequality (9) is the check of perfect correlations, for which we choose $\alpha_2 = \beta_3 = 0^\circ$, that

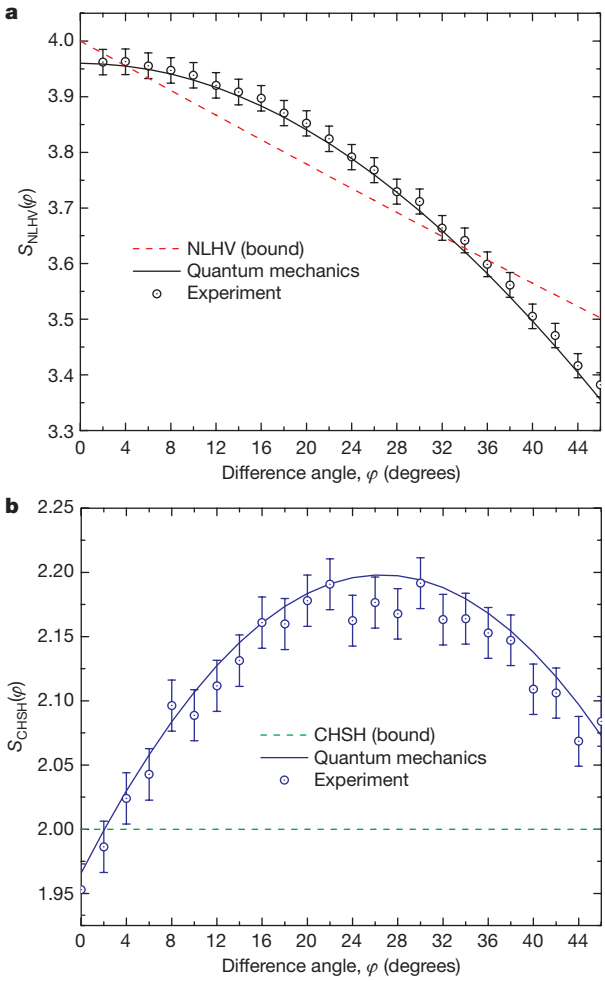


Figure 3 | Experimental violation of the inequalities for non-local hidden-variable theories (NLHV) and for local realistic theories (CHSH). a, Dashed line indicates the bound of inequality (9) for the investigated class of non-local hidden-variable theories (see text). The solid line is the quantum theoretical prediction reduced by the experimental visibility. The shown experimental data were taken for various difference angles φ (on the Poincaré sphere) of local measurement settings. The bound is clearly violated for $4^\circ < \varphi < 36^\circ$. Maximum violation is observed for $\varphi_{max} \approx 20^\circ$. **b,** At the same time, no local realistic theory can model the correlations for the investigated settings as the same set of data also violates the CHSH inequality (10). The bound (dashed line) is overcome for all values φ around φ_{max}, and hence excludes any local realistic explanation of the observed correlations in **a**. Again, the solid line is the quantum prediction for the observed experimental visibility. Error bars indicate s.d.

is, the intersection of the two orthogonal planes. Figure 3 shows the experimental violation of inequalities (9) and (10) for various difference angles. Maximum violation of inequality (9) is achieved, for example, for the settings $\{\alpha_1,\alpha_2,\beta_1,\beta_2,\beta_3\}=\{45^\circ,0^\circ,55^\circ,10^\circ,0^\circ\}$.

We finally obtain the following expectation values for the optimal settings for a test of inequality (9) (the errors are calculated assuming that the counts follow a poissonian distribution): $E(\mathbf{a}_1,\mathbf{b}_1) = -0.9298 \pm 0.0105$, $E(\mathbf{a}_2,\mathbf{b}_2) = -0.942 \pm 0.0112$, $E(\mathbf{a}_2,\mathbf{b}_3) = -0.9902 \pm 0.0118$. This results in $S_{\mathrm{NLHV}} = 3.8521 \pm 0.0227$, which violates inequality (9) by 3.2 standard deviations (see Fig. 3). At the same time, we can extract the additional correlation functions $E(\mathbf{a}_2,\mathbf{b}_1) = 0.3436 \pm 0.0088$, $E(\mathbf{a}_1,\mathbf{b}_2) = 0.0374 \pm 0.0091$ required for the CHSH inequality. We obtain $S_{\mathrm{CHSH}} = 2.178 \pm 0.0199$, which is a violation by ~9 standard deviations. The stronger violation of inequality (10) is due to the relaxed visibility requirements on the probed entangled state.

We have experimentally excluded a class of important non-local hidden-variable theories. In an attempt to model quantum correlations of entangled states, the theories under consideration assume realism, a source emitting classical mixtures of polarized particles (for which Malus' law is valid) and arbitrary non-local dependencies via the measurement devices. Besides their natural assumptions, the main appealing feature of these theories is that they allow us both to model perfect correlations of entangled states and to explain all existing Bell-type experiments. We believe that the experimental exclusion of this particular class indicates that any non-local extension of quantum theory has to be highly counterintuitive. For example, the concept of ensembles of particles carrying definite polarization could fail. Furthermore, one could consider the breakdown of other assumptions that are implicit in our reasoning leading to the inequality. These include Aristotelian logic, counterfactual definiteness, absence of actions into the past or a world that is not completely deterministic[29]. We believe that our results lend strong support to the view that any future extension of quantum theory that is in agreement with experiments must abandon certain features of realistic descriptions.

Received 22 December 2006; accepted 13 February 2007.

1. Schrödinger, E. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* **23**, 807–812; 823–828; 844–849 (1935).
2. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23,** 880–884 (1969).
3. Leggett, A. J. Nonlocal hidden-variable theories and quantum mechanics: An incompatibility theorem. *Found. Phys.* **33,** 1469–1493 (2003).
4. Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47,** 777–780 (1935).
5. Einstein, A. to Erwin Schrödinger, 19 June 1935 (Albert Einstein Archives, Jewish National and University Library, The Hebrew University of Jerusalem)
6. Bohr, N. in *Albert Einstein: Philosopher-Scientist* Vol. 7 (ed. Schilpp, P. A.) 201–241 (Library of Living Philosophers, Evanston, 1949).
7. Bell, J. S. On the Einstein Podolsky Rosen paradox. *Physics* **1,** 195–200 (1964).
8. Clauser, J. F. & Horne, M. D. Experimental consequences of objective local theories. *Phys. Rev. D* **10,** 526–535 (1974).
9. Clauser, J. F. in *Quantum [Un]speakables: From Bell to Quantum Information* (eds Bertlmann, R. A. & Zeilinger, A.) 61–98 (Springer, Heidelberg, 2002).
10. Greenberger, D. M., Horne, M. A. & Zeilinger, A. in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe* (ed. Kafatos, M.) 69–72 (Kluwer, Dordrecht, 1989).
11. Greenberger, D. M., Horne, M. A., Shimony, A. & Zeilinger, A. Bell's theorem without inequalities. *Am. J. Phys.* **58,** 1131–1143 (1990).
12. Freedman, S. J. & Clauser, J. F. Experimental test of local hidden-variable theories. *Phys. Rev. Lett.* **28,** 938–941 (1972).
13. Aspect, A., Grangier, P. & Roger, G. Experimental tests of realistic local theories via Bell's theorem. *Phys. Rev. Lett.* **47,** 460–463 (1981).
14. Aspect, A., Dalibard, J. & Roger, G. Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.* **49,** 1804–1807 (1982).
15. Weihs, G., Jennewein, T., Simon, C., Weinfurter, H. & Zeilinger, A. Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett.* **81,** 5039–5043 (1998).
16. Rowe, M. A. *et al.* Experimental violation of a Bell's inequality with efficient detection. *Nature* **409,** 791–794 (2001).
17. Pan, J.-W., Bouwmeester, D., Daniell, M., Weinfurter, H. & Zeilinger, A. Experimental test of quantum nonlocality in three-photon Greenberger-Horne-Zeilinger entanglement. *Nature* **403,** 515–519 (2000).
18. Zeilinger, A. Testing Bell's inequalities with periodic switching. *Phys. Lett. A* **118,** 1–2 (1986).
19. Aspect, A. Bell's inequality test: more ideal than ever. *Nature* **398,** 189–190 (1999).
20. Grangier, P. Count them all. *Nature* **409,** 774–775 (2001).
21. Born, M. & Wolf, E. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* 2nd edn (Pergamon, Oxford, 1964).
22. Bacon, D. & Toner, B. F. Bell inequalities with auxiliary communication. *Phys. Rev. Lett.* **90,** 157904 (2003).
23. Ne'eman, Y. The problems in quantum foundations in the light of gauge theories. *Found. Phys.* **16,** 361–377 (1986).
24. Bohm, D. A suggested interpretation of the quantum theory in terms of ''hidden'' variables. I and II. *Phys. Rev.* **85,** 166–193 (1952).
25. Dewdney, C., Holland, P. R. & Kyprianidis, A. A causal account of non-local Einstein-Podolsky-Rosen spin correlations. *J. Phys. Math. Gen.* **20,** 4717–4732 (1987).
26. Wu, C. S. & Shaknov, I. The angular correlation of scattered annihilation radiation. *Phys. Rev.* **77,** 136 (1950).
27. Kocher, C. A. & Commins, E. D. Polarization correlation of photons emitted in an atomic cascade. *Phys. Rev. Lett.* **18,** 575–577 (1967).
28. Clauser, J. F. & Shimony, A. Bell's theorem. Experimental tests and implications. *Rep. Prog. Phys.* **41,** 1881–1927 (1978).
29. Bell, J. S. Free variables and local causality. *Dialectica* **39,** 103–106 (1985).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to A. J. Leggett for stimulating this work and for discussions. We also thank D. Greenberger, M. A. Horne, T. Jennewein, J. Kofler, S. Malin and A. Shimony for their comments. T.P. is grateful for the hospitality of the IQOQI, Vienna. M.A. thanks L. Gohlike for his hospitality at the Seven Pines VIII. We acknowledge support from the Austrian Science Fund (FWF), the European Commission, the Austrian Exchange Service (ÖAD), the Foundation for Polish Science (FNP), the Polish Ministry of Higher Education and Science, the City of Vienna and the Foundational Questions Institute (FQXi).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A. (markus.aspelmeyer@quantum.at) or A.Z. (zeilinger-office@quantum.at).

LETTERS

A quantum scattering interferometer

Russell A. Hart[1], Xinye Xu[1]†, Ronald Legere[1]† & Kurt Gibble[1]

The collision of two ultracold atoms results in a quantum mechanical superposition of the two possible outcomes: each atom continues without scattering, and each atom scatters as an outgoing spherical wave with an *s*-wave phase shift. The magnitude of the *s*-wave phase shift depends very sensitively on the interaction between the atoms. Quantum scattering and the underlying phase shifts are vitally important in many areas of contemporary atomic physics, including Bose–Einstein condensates[1–5], degenerate Fermi gases[6–9], frequency shifts in atomic clocks[10–12] and magnetically tuned Feshbach resonances[13]. Precise experimental measurements of quantum scattering phase shifts have not been possible because the number of scattered atoms depends on the *s*-wave phase shifts as well as the atomic density, which cannot be measured precisely. Here we demonstrate a scattering experiment in which the quantum scattering phase shifts of individual atoms are detected using a novel atom interferometer. By performing an atomic clock measurement using only the scattered part of each atom's wavefunction, we precisely measure the difference of the *s*-wave phase shifts for the two clock states in a density-independent manner. Our method will enable direct and precise measurements of ultracold atom–atom interactions, and may be used to place stringent limits on the time variations of fundamental constants[14].

In our experiment, we 'juggle' atoms[15] in an atomic fountain clock by launching two gaseous clouds of caesium atoms upwards in rapid succession (cloud 1 first, followed by cloud 2) by laser-cooling them in a frame that moves upwards at 2.5–3.4 m s^{-1}. Gravity slows the atoms and, after cloud 1 reaches its apogee and begins to fall, the two clouds pass through one another and the atoms collide. For a short time delay between launches, of the order of $\Delta t = 10$ ms, the relative velocity of the atoms is about $v_r = g\Delta t = 10\,\mathrm{cm\,s^{-1}}$ (where g is the acceleration due to gravity); this corresponds to a collision energy of $E = mv_r^2/4 = 40\,\mu\mathrm{K} \times k_B$ (where m is the atomic mass and k_B is Boltzmann's constant). The collision energy is much greater than the atoms' temperature of 250–500 nK. We prepare the atoms in cloud 1 in a pure $|F,m\rangle$ hyperfine state (for example, $|4,4\rangle$) and those in cloud 2 in one of the clock states ($|3,0\rangle$). Both clouds pass through a microwave (clock) cavity, which puts the atoms in cloud 2 in a coherent superposition of the two clock states, $|3,0\rangle$ and $|4,0\rangle$. The phase of this coherence precesses at the caesium clock frequency, $\nu \approx 9.2$ GHz (ref. 12). When the two clouds collide, the atoms scatter as illustrated in Fig. 1. The *s*-wave part of the atomic wavefunction in each clock state, $|3,0\rangle$ or $|4,0\rangle$, scatters off the atoms in cloud 1, acquiring an *s*-wave phase shift, δ_3 or δ_4. After scattering, the atoms fall back through the microwave cavity, which converts the phase difference between the clock coherence and the microwave field into a population difference of the clock states, which we detect. This population difference for the unscattered part of each atom's wavefunction yields the usual transition probability for a clock as a function of microwave frequency, known as Ramsey fringes[12] (red diamonds in Fig. 2a). Here we instead detect only the scattered part of each atom's wavefunction, for which the phase of the coherence is shifted by the difference of the *s*-wave phase shifts, $\Phi = \delta_3 - \delta_4$ (blue circles in Fig. 2a). In this way, we use atomic-clock interferometry to directly observe the difference of the *s*-wave phase shifts. To demonstrate this technique, we scatter the caesium clock states off $|4,4\rangle$ at $v_r = 9.92\,\mathrm{cm\,s^{-1}}$ and measure $\Phi = -0.141(8)$ rad.

To select a clock atom that scatters, we use the Doppler shift and a narrow two-photon Raman transition[15–17]. In Fig. 3, we show velocity distributions of the vertical velocity component (v_z) of cloud 2, prepared in $|3,0\rangle$, when it collides with $|4,4\rangle$ atoms in cloud 1. For the data in Fig. 3, the microwave pulses to the clock cavity are disabled so that cloud 2 is not prepared in a coherent superposition of the clock states. The velocity-selective probe pulse transfers atoms from $|3,0\rangle$ to $|4,0\rangle$ with a bandwidth of 1.4 cm s^{-1}, and we detect the number of atoms in $F = 4$. Before the probe pulse, we push the atoms in $F = 4$ from the fountain with a laser beam tuned to excite $F = 4$ atoms[15,17]. We push the $F = 4$ atoms either early, before cloud 1 enters the clock cavity, or late, right after both clouds return downward through the clock cavity. Early clearing gives the 'no-collisions' signal in Fig. 3a (violet), and late clearing allows the two clouds to collide before we clear cloud 1, giving the 'collisions' signal in Fig. 3a (aqua). In the magnified Fig. 3b, the difference between the 'collisions' and 'no-collisions' curves between $v_z = -5\,\mathrm{cm\,s^{-1}}$ and $2\,\mathrm{cm\,s^{-1}}$ represents scattered atoms (Fig. 3c). For both curves in Fig. 3a and b, we subtract the small backgrounds in Fig. 3d, obtained by inhibiting the preparation of cloud 2 in $|3,0\rangle$. To observe the Ramsey fringes of scattered atoms (blue circles in Fig. 2), we fix the probe velocity at $v_z = 0$ in Fig. 3c, which corresponds to 90° scattering, enable the microwave clock pulses, and then scan the frequency of the microwave clock pulses. The contributions to the Ramsey fringes at $v_z = 0$ from the

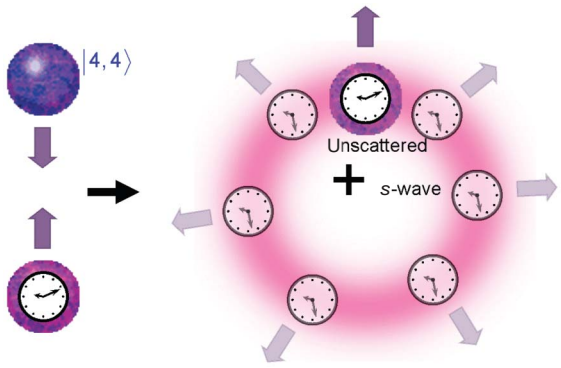


Figure 1 | Diagram of the experiment. Left, we collide an atom in a coherent superposition of the two caesium clock states (bottom) with a caesium atom in a pure $|F,m\rangle$ state, such as $|4,4\rangle$ (top). When the clock states scatter, they experience different *s*-wave phase shifts, shifting the phase of the clock coherence by the difference of the *s*-wave phase shifts. After the scattering (right), we directly observe the difference of the *s*-wave phase shifts by detecting only the scattered part of each clock atom's wavefunction.

[1]Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. †Present addresses: Department of Physics, East China Normal University, Shanghai 200062, China (X.X.); MIT Lincoln Laboratory, Lexington, Massachusetts 02420, USA (R.L.).

'no-collisions' and backgrounds are small, with phase shifts consistent with 0 (Fig. 2b and c). For this first demonstration, we choose 90° scattering to avoid contributions from *p*-waves.

These measurements are qualitatively different from the usual cold-collisions frequency shift in an operating atomic clock. The usual frequency shift arises from a collision rate which gives the rate at which the phase of the coherence is shifted. The usual frequency shift is due to the quantum-mechanical interference in the forward direction between the unscattered and scattered parts of each atom[18,19] and is proportional to density[10–12]. Here, the Ramsey fringes for the atoms that scatter have a phase shift (which is independent of density) instead of a frequency shift, because the phase of the coherence of the scattered part of each atom experiences both *s*-wave phase shifts, δ_3 and δ_4. To demonstrate the qualitative differences, we show in Fig. 4a the phase shift as a function of the free precession time T between the two microwave pulses for $T = 0.115$ s to 0.450 s. We increase the free precession time by increasing the launch velocity of both clouds, so that their apogees above the clock cavity are higher[12]. The phase shift is independent of T. It is clearly inconsistent with a frequency shift, for which the phase shift would increase linearly with T (dashed line in Fig. 4a). Further, the magnitudes of the effects are very different. The largest cold-collision frequency shift that has been observed in a clock is −5.5 mHz (ref. 10). Here, for $T = 0.115$ s, a phase shift of $\Phi = -0.141$ rad corresponds to a frequency shift of −200 mHz.

A key feature of this technique is that this difference of quantum scattering phase shifts is independent of the atomic density, to lowest order (Fig. 4b). Many experiments have probed a wide variety of scattering effects, including velocity redistribution[17,20], frequency shifts[10,11] and inelastic losses[21]. Each of these effects is given by a rate $nv_r\sigma$, where n is the atomic density and σ is the cross-section for the process. As the best measurements of the density of cold atoms do not achieve even 1% accuracy[22,23], it is generally not possible to precisely determine atomic scattering phase shifts and the atom–atom interactions from these measurements. Indeed for caesium, many scattering results appeared thoroughly inconsistent until the spectroscopic observation of many Feshbach resonances[24], combined with a theoretical analysis, determined the caesium interaction properties[25]. Here, although the number of scattered atoms and the amplitude of the Ramsey fringes are proportional to the density of cloud 1 (Fig. 4b inset), the phase shift is independent of the atomic density and therefore can utilize the high accuracy available with atomic clock techniques. In this first measurement, our statistical uncertainty is 6% (of even a relatively small difference of *s*-wave phase shifts), compared to typical density uncertainties of a factor of

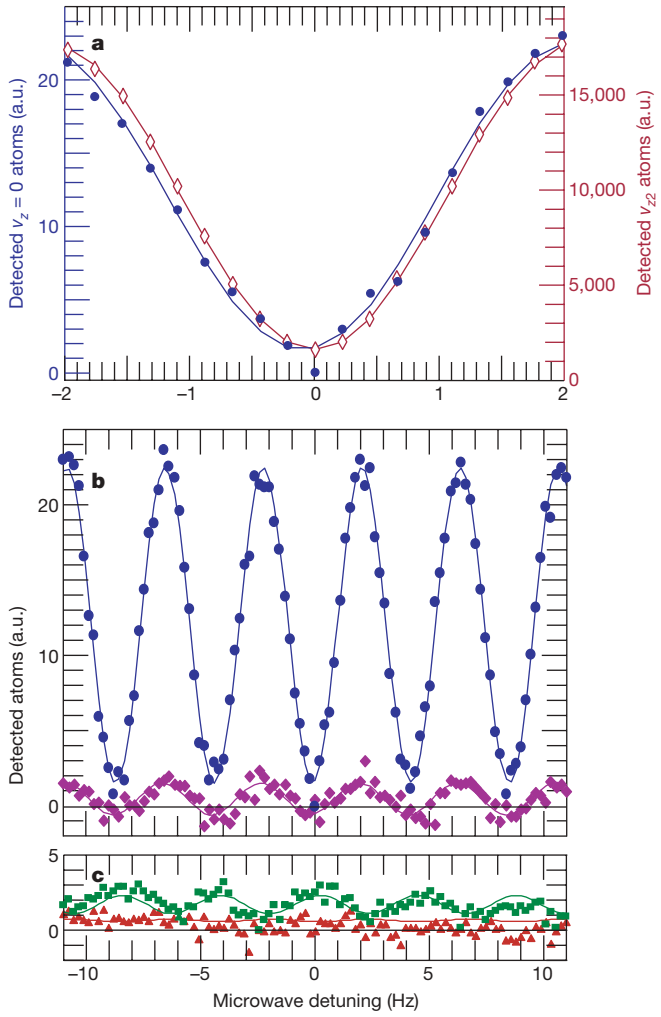


Figure 2 | Ramsey fringes for scattered and unscattered atoms. **a**, The central Ramsey fringe for clock atoms that have *s*-wave scattered from $|4,4\rangle$ atoms at 90° ($v_z = 0$, blue circles) with $v_r = 9.92\,\mathrm{cm\,s^{-1}}$. The reference Ramsey fringe is from the unscattered clock atoms (v_{z2}, maroon diamonds). The fringes for the scattered atoms have a phase shift of −0.141 rad, which is the difference of the *s*-wave phase shifts. a.u., arbitrary units. **b**, The entire Ramsey pattern for scattered atoms at $v_z = 0$ (blue circles) and for 'no-collisions' (violet diamonds) as in Fig. 3. **c**, The background Ramsey fringes as in Fig. 3 for 'collisions' (green squares) and 'no-collisions' (red triangles). For these data, the difference of the *s*-wave phase shifts for the clock states scattering off $|4,4\rangle$ is relatively small, $\Phi = -0.141(12)$ rad. Each circle represents the average of four differences of four measurements, requiring 16 cycles of the atomic fountain. The central Ramsey fringe is a minimum because the atoms begin in the $|3,0\rangle$ clock state and we detect the final number in $|3,0\rangle$. In **a**, there are nearly 1,000 times more unscattered atoms (maroon diamonds) than detected scattered atoms (blue circles).

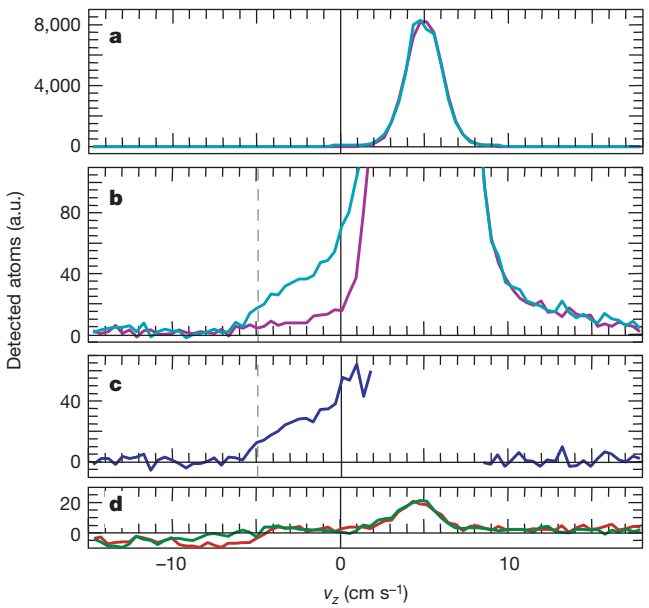


Figure 3 | Velocity distribution of clock atoms. Cloud 2 is prepared in $|3,0\rangle$ and cloud 1 in $|4,4\rangle$, with $v_r = 9.92\,\mathrm{cm\,s^{-1}}$. **a**, The aqua (violet) curves show the velocity distribution of cloud 2 for 'collisions' ('no-collisions') when we clear cloud 1 from the fountain late (early). **b**, Magnification of **a** by a factor of 100. In this centre-of-mass frame, the most probable vertical velocity component for cloud 1 is $v_{z1} = -4.96\,\mathrm{cm\,s^{-1}}$ (dashed line) and $v_{z2} = 4.96\,\mathrm{cm\,s^{-1}}$ for cloud 2. **c**, The difference between the 'collisions' and 'no-collisions' curves. This represents scattered atoms, which are visible between $v_z = -5\,\mathrm{cm\,s^{-1}}$ and $2\,\mathrm{cm\,s^{-1}}$. **d**, Background for 'collisions' (green) and 'no-collisions' (red). The Ramsey fringes in Fig. 2 (blue, violet, green, and red) are taken at $v_z = 0$ here on the curve of the same colour. The Ramsey fringes in Fig. 2 for unscattered atoms (maroon diamonds) are taken at v_{z2}. About 0.1% of the atoms scatter into the $1.4\,\mathrm{cm\,s^{-1}}$ detected velocity width at 90°.

two. Although measurements of differential cross-sections also yield density-independent phase shift differences[15,26,27], potential systematic errors that depend on the scattering angle limit the precision. Future improvements of our precision by orders of magnitude are expected.

At low energies, the atom–atom interactions are described by the *s*-wave scattering length, *a*. The scattering length is the low-temperature limit of $-\delta/k$ (where $k = \pi m v_r/h$ is the atomic wave vector and h is Planck's constant). Therefore, direct measurements of the difference of *s*-wave phase shifts can directly give precise differences of *s*-wave scattering lengths[5]. Using $\delta = -ka$, our current precision would translate to a scattering length difference uncertainty of ± 0.7 Å, comparable to the current uncertainty of the caesium triplet scattering length, $a = 1291.2(5)$ Å (ref. 25). A future accuracy of 100 μrad for Φ yields ± 0.009 Å, or 7 p.p.m. However, the caesium triplet scattering length is so large that $ka > 1$ for even $E = 1\ \mu\mathrm{K} \times k_B$. Therefore, our sensitivity to the caesium interatomic potentials at these energies is not so simple, and theoretical work is required to establish the sensitivity. Preliminary work has shown a sensitivity to scattering lengths of 1–100 p.p.m. for a measurement accuracy of 100 μrad (ref. 19). Chin and Flambaum[14] have recently suggested that highly sensitive measurements of scattering lengths near Feshbach resonances will set stringent limits on the time variation of the electron/proton mass ratio, a fundamental constant of physics. Near a Feshbach resonance, the phase shift in Fig. 2a has a resonant structure and varies by π. Our technique can accurately measure the phase shift throughout a resonance.

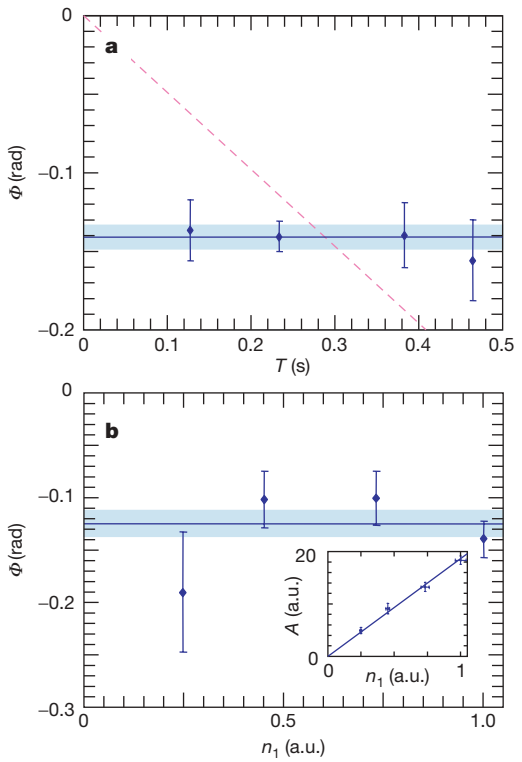


Figure 4 | The *s*-wave phase shift of scattered atoms. a, The phase shift Φ versus the free precession time T between the two microwave field interactions. The phase shift is independent of T, as opposed to being proportional to T as it is for a frequency shift (dashed line). We increase T by increasing the launch velocity in the fountain. The best fit to all the data is $\Phi = -0.141(8)$. **b**, The phase shift and Ramsey fringe amplitude, A, (inset) versus the atom density of cloud 1, n_1. The phase shift is independent of the density rather than proportional to it; in contrast, the usual cold-collision frequency shift in clocks is proportional to the density. As expected, the Ramsey fringe amplitude is proportional to the density of atoms in cloud 1. The error bars and bands represent standard errors.

We have demonstrated a fundamentally new scattering method that directly observes the phase shift of an atomic coherence due to quantum scattering. We use an atom interferometer in an atomic clock to accurately measure the difference of *s*-wave phase shifts. The technique is quite general; for any atom with a magnetic-field-insensitive transition, a variety of scattering channels can be explored with high accuracy. For caesium, any of the 16 hyperfine states can be studied as a function of collision energy and magnetic field for *s* waves, and for higher partial angular-momentum waves such as *p*, *d* and so on. The technique offers direct and unambiguous differences of quantum scattering phase shifts, and stringently probes ultracold atom–atom interactions.

METHODS

Juggling atomic fountain clock. Here we describe the changes to our juggling fountain, which is based on a double magneto-optic trap (MOT)[15,17]. We cool and optically pump each cloud of atoms after they are launched from the ultra-high-vacuum MOT in a three-dimensional moving-frame optical lattice with degenerate Raman-sideband cooling[28,29]. This optically pumps the atoms into the $|3,3\rangle$ state and cools them to a temperature of typically 500 nK for cloud 1, and 250 nK for cloud 2. Instead of launching both clouds with the same velocity, we launch cloud 1 with a slightly larger velocity than cloud 2 so that the two clouds collide after their apogees. The two clouds finish passing through one another just before they return downward through the microwave cavity. This has two advantages: (1) the two clouds are further separated at launch so that more of the atoms in cloud 1 survive the launch of cloud 2, and (2) it shortens the time between the collisions and the detection. After the clouds collide, the atoms spread out spherically with a velocity of $v_r/2 \approx 5\ \mathrm{cm\,s^{-1}}$. For the scattered atoms to be detected, they must pass through the 1.8 cm cavity apertures and be illuminated by the 2-cm-diameter detection laser beam. The highest collision rate occurs when the two cloud centres coincide, and the time from this point until we detect the scattered atoms is typically 0.13 s, so that the scattered cloud of these atoms spreads to a diameter of 1.3 cm.

We have added six 9.2-GHz microwave cavities to our juggling fountain to drive microwave transitions. The cavities used for the clock pulses and the state preparation are TE_{011} cylindrical cavities, dielectrically loaded with fused silica to reduce their size. The state-preparation cavities have 12-mm-diameter apertures in their endcaps, through which the atoms pass, and those for the clock cavity are 18 mm in diameter. Above the clock cavity, there is a $TE_{0,1,13}$ cavity that we use to probe the magnetic field[11], which is maintained near 15 mG. We pulse the microwaves to the clock cavity and, because of some leakage to the $TE_{0,1,13}$ cavity, we pulse the appropriate power and phase to it. We adjust the amplitude and phase so that the phase of the microwave field in the clock cavity is constant to less than 0.015 rad up to 7.4 mm above the centre of the clock cavity. In addition, we apply the 5-ms-long pulses when the atoms are centred in the clock cavity. In a future version, we will eliminate this leakage.

For the final state detection after the atoms pass downward through the clock cavity, we insert an aperture in the detection laser beam to limit the contributions from the unscattered atoms in clouds 1 and 2. With a vertical aperture height of the order of 1 cm, the background Ramsey fringes (violet, red and green curves in Fig. 2) are small. With no aperture, the background Ramsey fringes in Fig. 2 have amplitudes twice as large as the amplitude for scattered atoms, and still give the same phase shift.

State preparation. Both clouds of atoms are launched from the optical lattice in the $|3,3\rangle$ state and then pass through four microwave state-preparation cavities. A typical sequence for $T = 233$ ms follows (for other launch velocities, the order of the pulses may be slightly different). To minimize backgrounds, we purify the optical pumping of cloud 1 by transferring any residual atoms in $|3,1\rangle$ and $|3,2\rangle$ to the $F = 4$ hyperfine level with composite π pulses[30] in the first state-preparation cavity and pulsing a laser beam that pushes $F = 4$ atoms from the fountain. This is repeated in the second state-preparation cavity for $|3,-1\rangle$ and $|3,0\rangle$. After this, cloud 2 enters the first cavity and, as it travels through the first three cavities, a composite π pulse in each cavity transfers the atoms in $|3,3\rangle$ to $|4,3\rangle$ to $|3,2\rangle$, and finally to $|4,1\rangle$. In the middle of this sequence, cloud 1 is in the third cavity and we transfer those atoms from $|3,3\rangle$ to $|4,4\rangle$. After both clouds leave the cavities, we apply a laser pulse to repump any atoms left in $F = 3$. A stimulated-Raman transition then velocity-selectively transfers the atoms in cloud 2 from $|4,1\rangle$ to $|3,0\rangle$ before they enter the clock microwave cavity. The peak densities of cloud 1 (2) at launch are 6×10^9 $(1.2 \times 10^9)\ \mathrm{cm^{-3}}$, which we measure by observing collisions within one cloud[17] between $|3,0\rangle$ and $|4,0\rangle$, for which the triplet scattering length[25] gives a cross-section near the unitary limit. The number of atoms in cloud 1 and cloud 2 is about 1.6×10^9 and 3×10^8, respectively.

Received 5 January; accepted 8 February 2007.

1. Hall, D. S., Matthews, M. R., Ensher, J. R., Wieman, C. E. & Cornell, E. A. Dynamics of component separation in a binary mixture of Bose-Einstein condensates. *Phys. Rev. Lett.* **81,** 1539–1542 (1998).
2. Stenger, J. *et al.* Spin domains in ground-state Bose–Einstein condensates. *Nature* **396,** 345–348 (1998).
3. Khaykovich, L. *et al.* Formation of a matter-wave bright soliton. *Science* **296,** 1290–1293 (2002).
4. Strecker, K. E., Partridge, G. B., Truscott, A. G. & Hulet, R. G. Formation and propagation of matter-wave soliton trains. *Nature* **417,** 150–153 (2002).
5. Widera, A. *et al.* Precision measurement of spin-dependent interaction strengths for spin-1 and spin-2 ^{87}Rb atoms. *New J. Phys.* **8,** 152 (2006).
6. DeMarco, B. & Jin, D. S. Onset of Fermi degeneracy in a trapped atomic gas. *Science* **285,** 1703–1706 (1999).
7. Modugno, G. *et al.* Collapse of a degenerate Fermi gas. *Science* **297,** 2240–2243 (2002).
8. O'Hara, K. M., Hemmer, S. L., Gehm, M. E., Granade, S. R. & Thomas, J. E. Observation of a strongly interacting degenerate Fermi gas of atoms. *Science* **298,** 2179–2182 (2002).
9. Gupta, S. *et al.* Radio-frequency spectroscopy of ultracold fermions. *Science* **300,** 1723–1726 (2003).
10. Gibble, K. & Chu, S. Laser-cooled Cs frequency standard and a measurement of the frequency shift due to ultracold collisions. *Phys. Rev. Lett.* **70,** 1771–1774 (1993).
11. Fertig, C. & Gibble, K. Measurement and cancellation of the cold collision shift in an ^{87}Rb fountain clock. *Phys. Rev. Lett.* **85,** 1622–1625 (2000).
12. Wynands, R. & Weyers, S. Atomic fountain clocks. *Metrologia* **42,** S64–S79 (2005).
13. Inouye, S. *et al.* Observation of Feshbach resonances in a Bose–Einstein condensate. *Nature* **392,** 151–154 (1998).
14. Chin, C. & Flambaum, V. V. Enhanced sensitivity to fundamental constants in ultracold atomic and molecular systems near Feshbach resonances. *Phys. Rev. Lett.* **96,** 230801 (2006).
15. Legere, R. & Gibble, K. Quantum scattering in a juggling atomic fountain. *Phys. Rev. Lett.* **81,** 5780–5783 (1998).
16. Kasevich, M. *et al.* Atomic velocity selection using stimulated Raman transitions. *Phys. Rev. Lett.* **66,** 2297–2300 (1991).
17. Gibble, K., Chang, S. & Legere, R. Direct observation of s-wave atomic collisions. *Phys. Rev. Lett.* **75,** 2666–2669 (1995).
18. Legere, R. J. *Quantum Scattering in a Juggling Atomic Fountain*. PhD thesis, Yale Univ. (1999).
19. Kokkelmans, S. J. J. M. F. *Interacting Atoms in Clocks and Condensates*. PhD thesis, Univ. Eindhoven (2000).
20. Monroe, C. R., Cornell, E. A., Sackett, C. A., Myatt, C. J. & Wieman, C. E. Measurement of Cs-Cs elastic scattering at T = 30 μK. *Phys. Rev. Lett.* **70,** 414–417 (1993).
21. Myatt, C. J., Burt, E. A., Ghrist, R. W., Cornell, E. A. & Wieman, C. E. Production of two overlapping Bose-Einstein condensates by sympathetic cooling. *Phys. Rev. Lett.* **78,** 586–589 (1997).
22. Ensher, J. R., Jin, D. S., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Bose-Einstein condensation in a dilute gas: Measurement of energy and ground-state occupation. *Phys. Rev. Lett.* **77,** 4984–4987 (1996).
23. DePue, M. T., McCormick, C., Winoto, S. L., Oliver, S. & Weiss, D. S. Unity occupation of sites in a 3D optical lattice. *Phys. Rev. Lett.* **82,** 2262–2265 (1999).
24. Chin, C., Vuletic, V., Kerman, A. J. & Chu, S. High resolution Feshbach spectroscopy of cesium. *Phys. Rev. Lett.* **85,** 2717–2720 (2000).
25. Chin, C. *et al.* Precision Feshbach spectroscopy of ultracold Cs_2. *Phys. Rev. A* **70,** 032701 (2004).
26. Thomas, N. R., Kjærgaard, N., Julienne, P. S. & Wilson, A. C. Imaging of s and d partial-wave interference in quantum scattering of identical bosonic atoms. *Phys. Rev. Lett.* **93,** 173201 (2004).
27. Buggle, Ch, Léonard, J., von Klitzing, W. & Walraven, J. T. M. Interferometric determination of the s and d-wave scattering amplitudes in ^{87}Rb. *Phys. Rev. Lett.* **93,** 173202 (2004).
28. Hamann, S. E. *et al.* Resolved-sideband Raman cooling to the ground state of an optical lattice. *Phys. Rev. Lett.* **80,** 4149–4152 (1998).
29. Treutlein, P., Chung, K. Y. & Chu, S. High-brightness atom source for atomic fountains. *Phys. Rev. A* **63,** 051401(R) (2001).
30. Levitt, M. H. Composite pulses. *Prog. Nucl. Magn. Reson. Spectrosc.* **18,** 61–122 (1986).

Acknowledgements We acknowledge discussions with B. Verhaar and S. Kokkelmans and contributions from R. Li. This work was supported by NASA, NSF, ONR and Penn State University.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.G. (kgibble@psu.edu).

LETTERS

Helium isotopic evidence for episodic mantle melting and crustal growth

S. W. Parman[1]

The timing of formation of the Earth's continental crust is the subject of a long-standing debate[1,2], with models ranging from early formation with little subsequent growth, to pulsed growth, to steadily increasing growth. But most models do agree that the continental crust was extracted from the mantle by partial melting[3]. If so, such crustal extraction should have left a chemical fingerprint in the isotopic composition of the mantle. The subduction of oceanic crust and subsequent convective mixing, however, seems to have largely erased this record in most mantle isotopic systems (for example, strontium, neodymium and lead). In contrast, helium is not recycled into the mantle because it is volatile and degasses from erupted oceanic basalts. Therefore helium isotopes may potentially preserve a clearer record of mantle depletion than recycled isotopes. Here I show that the spectrum of ^{4}He/^{3}He ratios in ocean island basalts appears to preserve the mantle's depletion history, correlating closely with the ages of proposed continental growth pulses[4,5]. The correlation independently predicts both the dominant ^{4}He/^{3}He peak found in modern mid-ocean-ridge basalts, as well as estimates of the initial ^{4}He/^{3}He ratio of the Earth[6]. The correspondence between the ages of mantle depletion events and pulses of crustal production implies that the formation of the continental crust was indeed episodic and punctuated by large, potentially global, melting events. The proposed helium isotopic evolution model does not require a primitive, undegassed mantle reservoir, and therefore is consistent with whole mantle convection.

In the most widely cited He evolution models, the ^{4}He/^{3}He ratios observed in mid-ocean-ridge basalts (MORB) and ocean island basalts (OIB) are interpreted in terms of mixing between two reservoirs[7–10]. OIB are thought to sample an undegassed, low-^{4}He/^{3}He mantle that has retained most of its ^{3}He since the origin of the Earth and is often associated with the lower mantle. MORB are thought to sample a degassed, high-^{4}He/^{3}He mantle that has lost essentially all of its ^{3}He through melting[8,9,11] and is typically associated with the upper mantle. To maintain these two distinct reservoirs requires that they be chemically and physically isolated[7,11], and strongly implies some form of layered mantle convection, though alternative models can be envisaged[10].

However, this model rests upon the assumption that the lowest ^{4}He/^{3}He values found in OIB are representative of an undegassed mantle component, or at least place upper bounds on its value. This assumption is only valid if the partition coefficient for He (D^{He}, where $D^{He} = He_{crystal}/He_{melt}$) is lower than the partition coefficient for its parent isotopes, U and Th (D^{U+Th}). Recent experimental data suggest that D^{He} is in fact greater than D^{U+Th} (refs 12, 13). In this case, the mantle residues of melting ('depleted mantle') can preserve low ^{4}He/^{3}He ratios. Consistent with this, recent studies suggest that OIB and flood basalts with the lowest ^{4}He/^{3}He have depleted Sr, Nd and Pb isotopic compositions[14,15]. If this is so, then one cannot a priori assume that the lowest-^{4}He/^{3}He magmas are representative of an undegassed mantle, and in fact, there would be no requirement for undegassed mantle in the ^{4}He/^{3}He data.

Even though the experimental partitioning studies are ongoing and incomplete, there is enough evidence for high D^{He} at present for its potential consequences for He isotopic evolution models to be seriously considered. A fundamental difference from existing models is that if D^{He} is greater than D^{U+Th}, then the ^{4}He/^{3}He ratios in OIB could be interpreted in terms of an age of mantle melt depletion[16,17], rather than mixing between degassed and undegassed mantle reservoirs. This age would record the time of melting when U and Th were removed from the mantle residue to a greater extent than He, leaving it with a low (U+Th)/He (parent/daughter) ratio. It is essentially a U–Th depletion age. If D^{He} is more than about 5 times greater than D^{U+Th}, then the production of ^{4}He in the residue would effectively cease in the depleted residue and its ^{4}He/^{3}He ratio would remain constant thereafter.

MORB and OIB provide the primary constraints on the He isotopic composition of the present mantle. MORB generally have higher ^{4}He/^{3}He ratios than OIB (Fig. 1), with a dominant peak at $(90 \pm 1) \times 10^3$ (in terms of ^{3}He/^{4}He, a ratio of 7.9 R_a; R_a is the ^{3}He/^{4}He ratio of the atmosphere). There are peaks in the OIB data, but the importance of these has generally been downplayed. This is because a few islands (Hawaii, Iceland and Reunion) have a disproportionately large number of analyses relative to their actual volumes, and so they artificially dominate any data compilation. Even within single island groups, specific locations, such as Loihi seamount in Hawaii, are over-sampled with respect to their volume. However, a close inspection of the OIB data reveals that, when island groups are compared on an individual basis, there are certain ^{4}He/^{3}He peaks that recur in widely separated island groups (Fig. 1, Supplementary Fig. 1).

For most islands, there are only sufficient data to establish the existence of a few statistically significant peaks. However, Hawaii and Iceland both have over 400 analyses each, and so their ^{4}He/^{3}He spectrum can be resolved in greater detail. Most of their ^{4}He/^{3}He values lie in the range $(20–70) \times 10^3$. In this range, both islands have eight ^{4}He/^{3}He peaks and the positions of these peaks match quite well (Fig. 1). Other island groups show these same peaks, though any individual island group only contains a subset of them. In some cases, (Samoa, Kerguelen, Galapagos; see Supplementary Fig. 1), this may largely be due to lack of data, as their ^{4}He/^{3}He values roughly span the same range as Hawaii and Iceland. Some islands (Reunion, Azores, Canaries) clearly have a lower range of ^{4}He/^{3}He and fewer peaks. Above values of 70×10^3, the correlation between Hawaii, Iceland and other OIB largely ceases until the peak at 85×10^3.

This same gap, $(70–85) \times 10^3$, as well as the peaks at 50×10^3 and higher, can also be seen in the MORB data. The presence of the ^{4}He/^{3}He peaks at values below 90×10^3 in MORB are probably

[1]Department of Earth Sciences, University of Durham, South Road, Durham DH1 3LE, UK.

due to the influence of nearby plumes (the data plotted have not been filtered in any way). Plume-influenced ridges are generally shallow, and so can be filtered out of the data set by only considering MORB dredged from deep ridges (>3.5 km). When this is done, only data above 75×10^3 remain (Fig. 1). Thus the peaks in the MORB distribution lower than 75×10^3 could be transferred to the OIB distribution, but it would not change their position. If the $^4He/^3He$ peaks were purely due to over-sampling, then it is highly improbable that the same peaks would appear on different islands widely separated in distance. This suggests that whereas the relative heights of the peaks may largely be controlled by sampling choices and are not necessarily indicative of the $^4He/^3He$ spectrum of the mantle, the positions of the recurring peaks may have some underlying significance. In light of the experimental partitioning data, these peaks could represent times at which unusually large amounts of depleted mantle were produced, suggesting large temporal variations in global mantle melting rates.

One way to calculate ages for the $^4He/^3He$ peaks would be to assume a certain He evolution model and use it to obtain age estimates. However, there are a wide range of He evolution models that can be envisaged[10,18], and so a large range of non-unique depletion ages could be produced. Another approach is to look in the geologic record for independent estimates of the timing of mantle melting events, and in particular to see if the pattern of ages can be correlated with the pattern of OIB $^4He/^3He$ peaks. The continental crust may provide such a record of melting events. Studies of zircon U–Pb and Lu–Hf ages have suggested that there were a number of pulses of continental crust growth[4,5,19]. The main zircon age peaks are at 1.2, 1.9, 2.7 and 3.3 Gyr (Fig. 2). As the continental crust was extracted from the mantle by partial melting[3], these peaks could record pulses of mantle melting. Alternatively, they could merely represent areas of continental crust that have randomly escaped tectonic recycling and may not imply increased continental crust growth rates[2,20]. In this case, there should be no correlation between mantle depletion events and zircon age peaks.

The pattern of the zircon age peaks is distinctive, with a number of peaks in the Archaean and a large gap after 1.2 Gyr (Fig. 2). This is reminiscent of the pattern of the He peaks, where there are a number of peaks at low $^4He/^3He$ and then a large gap above 70×10^3. If one correlates the patterns on this basis, the points produce a line with a very high R^2 (0.9986, see Supplementary Table 1 for data points) on a He evolution diagram (from here on referred to as the He-continental crust (He-CC) correlation). Although other choices of how to correlate the peaks will produce linear correlations (because both the He and zircon age data sets are monotonically increasing), they have lower correlation coefficients.

More significantly, the slope of the line is not random. If one projects the He-CC correlation to the present day, it yields a value of $^4He/^3He$ of $(92.2 \pm 0.7) \times 10^3$ (±s.e.), which matches the value of the dominant peak in modern MORB of $(91 \pm 1.5) \times 10^3$ (error is width of peak, Fig. 1). Likewise, projected back to 4.55 Gyr ago, the age of the Earth, it yields a value of $(5.6 \pm 0.8) \times 10^3$ (±s.e.), matching the value estimated for the initial Earth from the atmosphere of Jupiter, $(6 \pm 0.2) \times 10^3$ (ref. 6). The independent prediction of these two fundamental values by the He-CC correlation is strong evidence that the He and zircon peaks are causally related. The slope is a robust feature of the data and is not controlled by any one point. For instance, if only the four main peaks (1.2, 1.9, 2.7 and 3.3 Gyr) are used, the R^2 is 0.9994, the $t = 0$ intercept is $(91.1 \pm 0.8) \times 10^3$ and the $t = 4.55$ Gyr intercept is $(7.3 \pm 0.7) \times 10^3$. If the He-CC correlation was spurious, or the He peaks themselves were simply the result of over-sampling certain locations, it would require three highly improbable coincidences: (1) the correlation of $^4He/^3He$ peaks from island

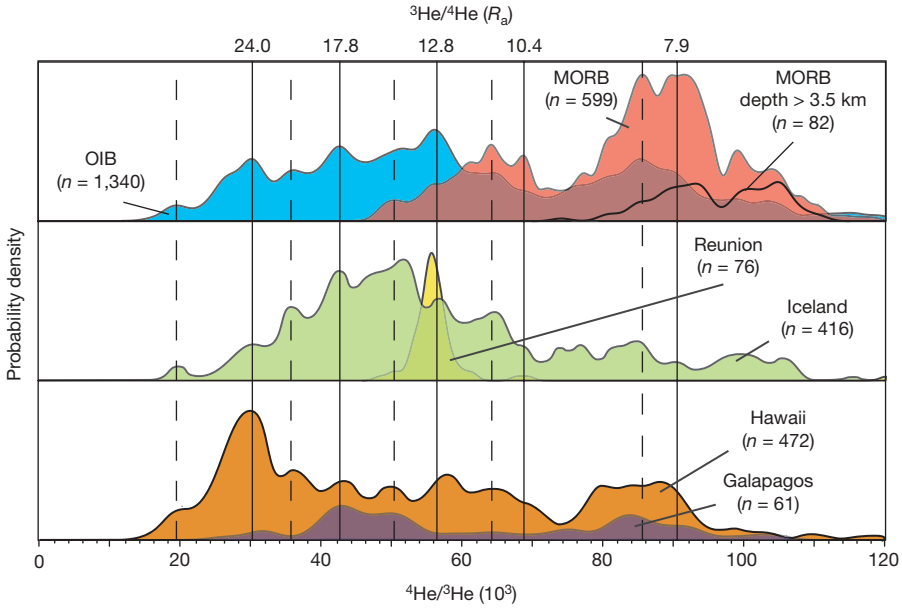


Figure 1 | Globally recurring $^4He/^3He$ peaks in oceanic basalts. Curves show probability density functions (PDFs, n = number of data points in the distribution) of $^4He/^3He$ ratios in ocean island basalts (OIB: blue), mid-ocean-ridge basalts (MORB: red, unfiltered; dredged from depths >3.5 km, black line), Iceland (green), Reunion (yellow), Hawaii (orange) and Galapagos (purple). Data are taken from ref. 31. Vertical lines show peaks that are present in more than one ocean island (see Supplementary Fig. 1 for histograms of the data as well as distributions for other islands) and so are unlikely to be sampling bias artefacts. Solid vertical lines are for peaks that correspond to the four major zircon age peaks[4,5] (see Fig. 2). Dashed lines correspond to smaller zircon age peaks, except for the peak at 20×10^3, which has no matching zircon peak. The $^3He/^4He$ values for the peaks are given along the top, in units of R_a (R_a is the $^3He/^4He$ ratio of the atmosphere). The MORB database has not been filtered, so samples with $^4He/^3He$ lower than 75×10^3 are probably attributable to the influence of nearby plumes, because the peaks are not present when the data are filtered for depth (see Supplementary Fig. 1). Thus those data could be transferred to the OIB PDF, but the peak positions would not be affected. The zircon age peaks in the range $(20–70) \times 10^3$ are well expressed in the OIB $^4He/^3He$ data, but less so after that. The 69×10^3 peak is barely visible as a shoulder on the Hawaii and Iceland data, and is perhaps represented by two points in the Reunion data that make a small peak. It is much clearer in the MORB data. The peak at 86×10^3 in the OIB data is largely due to Atlantic islands (Azores and Cape Verdes), though it is present, if less sharp, in Pacific islands (Galapagos) as well.

to island, (2) the correspondence of those peaks with the zircon age peak pattern and (3) the prediction of both the present-day MORB and initial Earth $^4He/^3He$ values.

If the He-CC correlation is significant, it implies that the crustal growth has been punctuated by periods of unusually high growth rates[4,5,19], and that these pulses are recorded in both the mantle (as He peaks in OIB and MORB) and the crust (as zircon age peaks). These events would be catastrophic (as opposed to uniformitarian), and it is not clear that there are any Phanerozoic analogues. Interestingly, komatiites have a similar age distribution as the melting pulses, with most being erupted at 3.3, 2.7 and 1.9 Gyr ago. Perhaps their unique geochemistry provides clues to the nature of the large melting events.

A clear requirement of the He-CC correlation is that depleted mantle domains have been preserved for billions of years against mantle convection. How this happens is not clear, and neither is the spatial distribution of the domains within the mantle. A number of papers have pointed out that many ocean islands are spatially associated with seismic low-velocity areas at the core–mantle boundary, and that these regions may be the ultimate source of OIB[21–23]. Such a dense boundary layer would be a good place to preserve depleted heterogeneities against convective rehomogenization for long time periods.

The linear form of the He-CC correlation is a strong observational constraint on any He evolution model. A wide range of He models have been proposed, many of which produce linear $^4He/^3He$ evolution owing to the degassing of the mantle[10,18]. How much of the mantle has been degassed is not constrained by the He peaks, but as the entire spectrum of $^4He/^3He$ values (below 91×10^3) is interpreted to reflect ancient, depleted mantle domains, there is no requirement (or evidence) for an undepleted mantle reservoir or for layered convection in this interpretation of He isotope systematics.

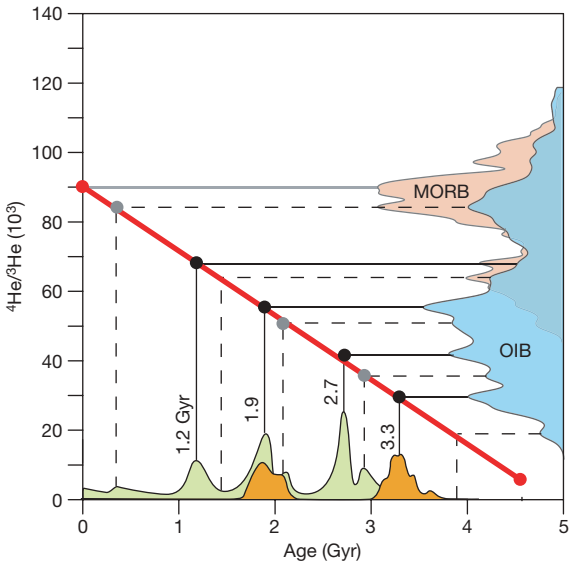


Figure 2 | Correspondence of OIB and MORB $^4He/^3He$ peaks with continental crust zircon age peaks. Probability distribution functions of OIB (blue) and MORB (red) from Fig. 1 are shown along the right side of the figure. Crustal zircon age distributions are shown along the horizontal axis (orange[5], green[4] fields). There are seven He peaks in the OIB and MORB distributions that can be correlated to the zircon age peaks (connected by solid and dashed lines). A regression through these points (black, major peaks; grey, minor peaks) yields an R^2 of 0.9986, an intercept at $t = 0$ of $(92.2 \pm 0.7) \times 10^3$ and an intercept at 4.56 Gyr of $(5.6 \pm 0.8) \times 10^3$ (thick red line). Thus it independently predicts the dominant peak in the MORB data ($(91.0 \pm 1.5) \times 10^3$, filled red circle at age = 0 Gyr) and the initial $^4He/^3He$ of the Earth ($(6.0 \pm 0.2) \times 10^3$, estimated from the atmosphere of Jupiter[6], red filled circle at age = 4.55 Gyr). The thick red line represents the He isotopic evolution of the MORB source. The $^4He/^3He$ evolution is linear with time because He is not recycled back into the mantle by subduction, and so the mantle is an open system with respect to He (see text). Two He peaks do not have matching zircon peaks, corresponding to ages of 3.9 and 1.4 Gyr. Thus one prediction of the He-CC correlation is that zircon age peaks might be found at these ages.

If there is no primitive mantle, it implies that the entire mantle has been processed by melting and consists mostly of depleted, harzburgitic material with a subordinate amount of recycled oceanic crust. Such a view has been previously proposed and is consistent with seismic scattering observations[24–27]. If this is the case, then most of the incompatible trace elements in the mantle are in the volumetrically minor eclogitic components, whereas most of the compatible elements are in the harzburgitic matrix. Helium is unusual in that it is an incompatible element that is not enriched in the recycled component because it degasses from erupted lavas and, from there, escapes to space. Thus both depleted and recycled components may have roughly similar He concentrations. Indeed, He concentrations may be higher in the ancient depleted domains as they seem to have acted as closed systems for much of Earth's history, whereas the recycled components may have undergone multiple cycles of degassing. In either case, harzburgitic material makes up >70% of the mantle in this view and so would contain the bulk of the Earth's He. This may explain the apparently paradoxical situation that OIB are enriched in nearly all incompatible trace elements and isotopic systems[28] (reflecting the contribution from the recycled components), but have depleted He isotopes (reflecting the presence of ancient depleted harzburgite material that makes a negligible contribution to the trace element budget of the magmas but dominates the He budget).

The ultimate cause of these large melting events is not clear. Presumably, they are related to large releases of heat from the mantle (though a large flux of volatiles could also produce melting). Numerical mantle convection models have produced large variations in mantle heat flux owing to variable rates of mantle convection[29,30]. Perhaps the large melting events record such large-scale mantle overturns.

Received 4 October 2006; accepted 16 February 2007.

1. Taylor, S. R. & McLennan, S. M. The geochemical evolution of the continental crust. *Rev. Geophys.* **33,** 241–265 (1995).
2. Bowring, S. A. & Housh, T. The Earth's early evolution. *Science* **269,** 1535–1540 (1995).
3. Hofmann, A. W. Chemical differentiation of the Earth — the relationship between mantle, continental crust, and oceanic crust. *Earth Planet. Sci. Lett.* **90,** 297–314 (1988).
4. Condie, K. C. Episodic continental growth and supercontinents: a mantle avalanche connection? *Earth Planet. Sci. Lett.* **163,** 97–108 (1998).
5. Kemp, A. I. S., Hawkesworth, C. J., Paterson, B. A. & Kinny, P. D. Episodic growth of the Gondwana supercontinent from hafnium and oxygen isotopes in zircon. *Nature* **439,** 580–583 (2006).
6. Mahaffy, P. R., Donahue, T. M., Atreya, S. K., Owen, T. C. & Niemann, H. B. Galileo probe measurements of D/H and $^3He/^4He$ in Jupiter's atmosphere. *Space Sci. Rev.* **84,** 251–263 (1998).
7. Allègre, C. J. & Moreira, M. Rare gas systematics and the origin of oceanic islands: the key role of entrainment at the 670 km boundary layer. *Earth Planet. Sci. Lett.* **228,** 85–92 (2004).
8. Kurz, M. D., Jenkins, W. J. & Hart, S. R. Helium isotopic systematics of oceanic islands and mantle heterogeneity. *Nature* **297,** 43–47 (1982).
9. Moreira, M., Breddam, K., Curtice, J. & Kurz, M. D. Solar neon in the Icelandic mantle: new evidence for an undegassed lower mantle. *Earth Planet. Sci. Lett.* **185,** 15–23 (2001).
10. Porcelli, D. & Ballentine, C. J. Models for the distribution of terrestrial noble gases and evolution of the atmosphere. *Rev. Mineral. Geochem.* **47,** 411–480 (2002).
11. Allègre, C. J., Hofmann, A. & O'Nions, K. The argon constraints on mantle structure. *Geophys. Res. Lett.* **23,** 3555–3557 (1996).
12. Brooker, R. A. *et al.* The 'zero charge' partitioning behaviour of noble gases during mantle melting. *Nature* **423,** 738–741 (2003).
13. Parman, S. W., Kurz, M. D., Hart, S. R. & Grove, T. L. Helium solubility in olivine and implications for high $^3He/^4He$ in ocean island basalts. *Nature* **437,** 1140–1143 (2005).
14. Class, C. & Goldstein, S. L. Evolution of helium isotopes in the Earth's mantle. *Nature* **436,** 1107–1112 (2005).
15. Stuart, F. M., Lass-Evans, S., Fitton, J. G. & Ellam, R. M. High $^3He/^4He$ ratios in picritic basalts from Baffin Island and the role of a mixed reservoir in mantle plumes. *Nature* **424,** 57–59 (2003).

16. Graham, D., Lupton, J., Albarède, F. & Condomines, M. Extreme temporal homogeneity of helium isotopes at Piton de la Fournaise, Réunion Island. *Nature* **347,** 545–548 (1990).
17. Kurz, M. D. Mantle heterogeneity beneath oceanic islands — some inferences from isotopes. *Phil. Trans. R. Soc. Lond.* **342,** 91–103 (1993).
18. Seta, A., Matsumoto, T. & Matsuda, J. Concurrent evolution of $^{3}He/^{4}He$ ratio in the Earth's mantle reservoirs for the first 2 Ga. *Earth Planet. Sci. Lett.* **188,** 211–219 (2001).
19. Stein, M. & Hofmann, A. W. Mantle plumes and episodic crustal growth. *Nature* **372,** 63–68 (1994).
20. Armstrong, R. L. Radiogenic isotopes — The case for crustal recycling on a near-steady-state no-continental-growth Earth. *Phil. Trans. R. Soc. Lond.* **301,** 443–472 (1981).
21. Burke, K. & Torsvik, T. H. Derivation of large igneous provinces of the past 200 million years from long-term heterogeneities in the deep mantle. *Earth Planet. Sci. Lett.* **227,** 531–538 (2004).
22. Courtillot, V., Davaille, A., Besse, J. & Stock, J. Three distinct types of hotspots in the Earth's mantle. *Earth Planet. Sci. Lett.* **205,** 295–308 (2003).
23. Wen, L. X. A compositional anomaly at the Earth's core-mantle boundary as an anchor to the relatively slowly moving surface hotspots and as source to the DUPAL anomaly. *Earth Planet. Sci. Lett.* **246,** 138–148 (2006).
24. Allègre, C. J. & Turcotte, D. L. Implications of a two-component marble-cake mantle. *Nature* **323,** 123–127 (1986).
25. Helffrich, G. R. & Wood, B. J. The Earth's mantle. *Nature* **412,** 501–507 (2001).
26. Hofmann, A. W. & White, W. M. Mantle plumes from ancient oceanic crust. *Earth Planet. Sci. Lett.* **57,** 421–436 (1982).
27. Davies, G. F. Stirring geochemistry in mantle convection models with stiff plates and slabs. *Geochim. Cosmochim. Acta* **66,** 3125–3142 (2002).
28. Zindler, A. & Hart, S. Chemical geodynamics. *Annu. Rev. Earth Planet. Sci.* **14,** 493–571 (1986).
29. Butler, S. L., Peltier, W. R. & Costin, S. O. Numerical models of the Earth's thermal history: Effects of inner-core solidification and core potassium. *Phys. Earth Planet. Inter.* **152,** 22–42 (2005).
30. Davies, G. F. Punctuated tectonic evolution of the earth. *Earth Planet. Sci. Lett.* **136,** 363–379 (1995).
31. Abedini, A. A., Hurwitz, S. & Evans, W. C. USGS-NoGaDat — A global dataset of noble gas concentrations and their isotopic ratios in volcanic systems. (US Geological Survey Digital Data Series, 202) ⟨http://pubs.usgs.gov/ds/2006/202⟩ (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I thank G. Pearson, P. Martin and T. Grove for informal reviews, M. Kurz, S. Hart, M. Walter and J. Blundy for discussions, and A. Abedini for compiling the He isotope database. This research was supported by the Nuffield Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to the author (stephen.parman@durham.ac.uk).

ARTICLES

Anaphase initiation is regulated by antagonistic ubiquitination and deubiquitination activities

Frank Stegmeier[1]*, Michael Rape[3]*†, Viji M. Draviam[3,4], Grzegorz Nalepa[2], Mathew E. Sowa[2], Xiaolu L. Ang[2], E. Robert McDonald III[1], Mamie Z. Li[1], Gregory J. Hannon[5], Peter K. Sorger[3,4], Marc W. Kirschner[3], J. Wade Harper[2] & Stephen J. Elledge[1]

The spindle checkpoint prevents chromosome mis-segregation by delaying sister chromatid separation until all chromosomes have achieved bipolar attachment to the mitotic spindle. Its operation is essential for accurate chromosome segregation, whereas its dysregulation can contribute to birth defects and tumorigenesis. The target of the spindle checkpoint is the anaphase-promoting complex (APC), a ubiquitin ligase that promotes sister chromatid separation and progression to anaphase. Using a short hairpin RNA screen targeting components of the ubiquitin-proteasome pathway in human cells, we identified the deubiquitinating enzyme USP44 (ubiquitin-specific protease 44) as a critical regulator of the spindle checkpoint. USP44 is not required for the initial recognition of unattached kinetochores and the subsequent recruitment of checkpoint components. Instead, it prevents the premature activation of the APC by stabilizing the APC-inhibitory Mad2–Cdc20 complex. USP44 deubiquitinates the APC coactivator Cdc20 both *in vitro* and *in vivo*, and thereby directly counteracts the APC-driven disassembly of Mad2–Cdc20 complexes (discussed in an accompanying paper). Our findings suggest that a dynamic balance of ubiquitination by the APC and deubiquitination by USP44 contributes to the generation of the switch-like transition controlling anaphase entry, analogous to the way that phosphorylation and dephosphorylation of Cdk1 by Wee1 and Cdc25 controls entry into mitosis.

Aneuploidy, a state of abnormal chromosome number, is a hallmark of aggressive tumours and can arise from errors during chromosome segregation[1–3]. Faithful chromosome segregation requires that each chromatid pair achieves bipolar attachment to the mitotic spindle before their segregation. Segregation is initiated by activation of the APC, an E3 ubiquitin ligase. When bound to its cofactor Cdc20, APC^{Cdc20} drives cells into anaphase by inducing the degradation of securin and mitotic cyclins[4]. The spindle checkpoint ensures that activation of the APC is delayed until all chromosomes have achieved bipolar kinetochore–microtubule attachment. Many checkpoint proteins are recruited to unattached kinetochores and are thought to facilitate the binding of Mad2 and BubR1 to Cdc20, thereby inhibiting APC^{Cdc20} activity[5–7].

The first spindle checkpoint components were discovered through genetic screens in budding yeast[8–10], and characterization of their metazoan orthologues showed the conservation of fundamental aspects of mitotic checkpoint signalling from yeast to human[3,6]. However, recent studies revealed that spindle checkpoint signalling is more complex in mammalian cells, as Zw10 (zeste-white 10), Rod (rough-deal) and $p31^{comet}$ regulate spindle checkpoint function in higher eukaryotes but lack clear yeast orthologues[11–15].

Here we use a genetic approach to identify novel regulators of spindle checkpoint control in human cells. By screening a short hairpin (sh)RNA library that targets genes in the ubiquitin-proteasome pathway, we have identified a previously uncharacterized deubiquitinating enzyme, USP44. We find that USP44 acts downstream of kinetochore-localized spindle checkpoint complexes and prevents premature activation of the APC by stabilizing the APC-inhibitory Mad2–Cdc20 complex. USP44 can directly deubiquitinate Cdc20 and counteract the APC-driven disassembly of spindle checkpoint complexes[16] both *in vitro* and *in vivo*. Our study indicates that spindle checkpoint regulation of the APC is based on a dynamic balance of counteracting ubiquitination and deubiquitination activities.

shRNA screen for spindle checkpoint regulators

Given that the central target of the spindle checkpoint is the APC[4–7], we wished to explore possible roles of other components of the ubiquitin pathway in the regulation of anaphase entry through the spindle checkpoint. Using a human microRNA-based genome-wide shRNA library as a source of hairpins[17], we developed an arrayed shRNA library containing 1,964 hairpins targeting 759 genes related to the ubiquitin pathway (see Supplementary Fig. 1b and Supplementary Table 1 for detailed library information). We screened this library for regulators of the spindle checkpoint using a high-throughput visual screening platform (Supplementary Fig. 1a–c). This screen was intended to identify two classes of genes whose depletion leads to a reduced mitotic index: (1) spindle checkpoint components, the inactivation of which by-passes Taxol-induced mitotic arrest; and (2) pre-mitotic cell-cycle regulators, the inactivation of which delays cells before mitotic entry. In contrast to the multi-lobed nuclear

[1]Howard Hughes Medical Institute, Department of Genetics, Harvard Partners Center for Genetics and Genomics, and [2]Department of Pathology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [3]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. †Present address: Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720-3202, USA.
*These authors contributed equally to this work.

morphology caused by checkpoint by-pass, cells that arrest before mitosis contained normal (unlobed) interphase nuclei, which allowed us to classify preliminarily our candidate genes into potential checkpoint regulators (MN for multi-lobed nuclei) and cell division cycle mutants (CDC, borrowing from the yeast nomenclature) (Supplementary Fig. 1d). Our screen identified known regulators of pre-mitotic cell-cycle progression (for example, *EMI1* (also called *FBXO5*) and *SKP2*) and 14 new candidate spindle checkpoint regulators and CDC mutants (Supplementary Fig. 1d). We validated the most penetrant candidate genes with additional siRNAs that target different regions within each gene (Fig. 1a, b and Supplementary Fig. 2), strengthening the notion that phenotypes are due to target down-regulation rather than off-target effects.

We focused our investigation on the candidate checkpoint protein USP44, a previously uncharacterized deubiquitinating enzyme, because it had the most penetrant phenotype of all candidate spindle checkpoint regulators. Depletion of USP44 abolished checkpoint function in all three cell lines tested (Fig. 1 and Supplementary Fig. 3), and the efficiency in the reduction of USP44 levels for different siRNAs strongly correlated with their phenotypic penetrance in the checkpoint assay (Fig. 1a, d). Furthermore, expression of an siRNA-resistant version of USP44 rescued the spindle checkpoint defect of USP44-depleted cells (Fig. 1e, USP44(siRes)), confirming the specificity of our RNAi constructs. In contrast, a catalytically inactive USP44 mutant failed to restore spindle checkpoint function (Fig. 1e, USP44(ci-siRes)). Thus, deubiquitination mediated by USP44 is essential for spindle checkpoint function in human cells.

USP44 activity is elevated in mitosis

Consistent with an essential function of USP44 in the spindle checkpoint, USP44 levels were increased in mitotic cells (Fig. 2a), but rapidly decreased once cells had completed chromosome attachment and exited from mitosis (Fig. 2b). We noted the appearance of a slower migrating form of USP44 as cells entered mitosis (Fig. 2a) that disappeared shortly after release from nocodazole-mediated arrest and preceded USP44's degradation (Fig. 2b). This slower migrating form was also evident when recombinant USP44 was incubated in mitotic extracts, and became more pronounced after treatment with the phosphatase inhibitor okadaic acid, indicating that USP44 may be phosphorylated in mitosis (Fig. 2c).

To test whether the deubiquitination activity of USP44 is cell-cycle regulated, we took advantage of the fact that USP44 purified from insect cells has very low catalytic activity in an *in vitro* deubiquitination assay (Fig. 2d). Notably, pre-incubation of recombinant USP44 with extracts from nocodazole-arrested cells strongly increased the activity of USP44, whereas treatment with S-phase extracts resulted only in residual activation similar to a catalytically inactive mutant (Fig. 2d). Together, these findings indicate that USP44 is activated in checkpoint-arrested mitotic cells and rapidly degraded as cells exit from mitosis.

USP44 prevents premature anaphase onset

To characterize the mitotic function of USP44 in an unperturbed mitosis, we monitored mitotic progression and chromosome movement in USP44-depleted HeLa cells expressing histone H2B tagged with green fluorescent protein (GFP). Downregulation of USP44 led to a variety of mitotic defects, including the failure to form a clear

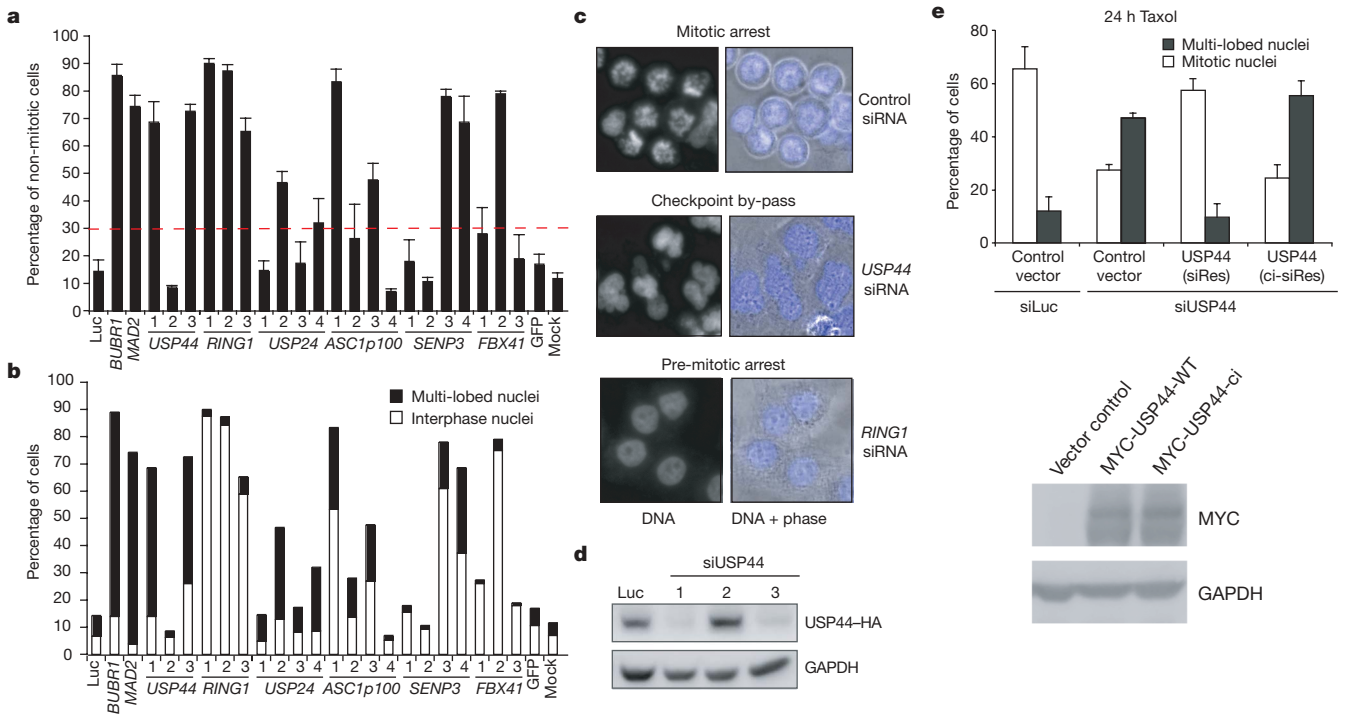


Figure 1 | Validation of candidate genes from Taxol screen. a, **b**, HeLa cells were transfected with siRNAs, treated with 100 nM Taxol 48 h after transfection, and fixed for visual inspection 24 h after Taxol addition. The percentage of non-mitotic cells (**a**) and cells with multi-lobed or interphase nuclei (**b**) was quantified. The values represent averages of three independent experiments ($n = 100$, error bars ± 1 s.d.). The horizontal dashed line indicates the applied threshold for Taxol screen (30% non-mitotic cells). **c**, Representative images of HeLa cells transfected with siRNAs targeting luciferase (control), *USP44* (oligo1), or *RING1* (oligo1) that illustrate the phenotypic differences between mitotic arrest, checkpoint by-pass and pre-mitotic arrest. **d**, HeLa cells stably expressing haemagglutinin (HA)-tagged USP44 were transfected with siRNAs, cell extracts collected 48 h after transfection, and probed with the indicated antibodies. **e**, Transfection of siRNA-resistant wild-type Myc-tagged USP44 (USP44(siRes)), but not the catalytically inactive-mutant USP44 (USP44(ci-siRes)), rescues the spindle checkpoint defect of cells treated with *USP44*-targetted siRNA. HeLa cells were transfected with the indicated siRNAs and rescuing plasmids and treated as described for Fig. 1a. Western blot shows similar expression levels of wild-type and mutant Myc–USP44 ($n = 100$; error bars indicate ± 1 s.d.).

metaphase plate (fourfold increase), unaligned and lagging chromosomes during anaphase (fivefold and 15-fold increase, respectively), and cytokinesis defects (sevenfold increase) (Fig. 2e, f and Supplementary Fig. 4a). Thus, USP44 has an essential function in an unperturbed cell cycle like other checkpoint proteins[5,6].

When we compared the timing of anaphase onset in cells with reduced USP44, Mad1 or Mad2 function by time-lapse imaging, we found that USP44 depletion markedly accelerated the onset of anaphase, as occurs with reduced levels of Mad2 but not Mad1 (Fig. 2g)[18]. Consistent with accelerated mitotic progression, USP44-depleted cells degraded cyclin B prematurely in prometaphase (Fig. 3a, b), whereas the timing of cyclin A degradation (which is not responsive to checkpoint activation *in vivo*[19,20]) was unaffected (Fig. 3c). These findings, together with the fact that the kinetics of checkpoint by-pass in response to nocodazole treatment were similar in Mad2- and USP44-depleted cells (Supplementary Fig. 4b), indicated that USP44 may control spindle checkpoint function by regulating Mad2.

USP44 regulates Mad2–Cdc20 association

The order of function of spindle checkpoint proteins has been determined by examining their interdependencies of kinetochore recruitment. Currently, all known checkpoint proteins are required for Mad2 localization, placing them upstream of Mad2. By that criterion, USP44 must be downstream of or parallel to Mad2, because Mad2, as well as Bub1, BubR1, Mad1 and Mps1, were all properly localized in USP44-depleted prometaphase cells (Fig. 3d, e and Supplementary Fig. 5).

These results suggest that despite USP44 depletion, cells can detect unattached kinetochores and properly recruit spindle checkpoint proteins. Yet these same cells apparently fail to restrain APC^{Cdc20} activity. APC^{Cdc20} inhibition by the spindle checkpoint is mediated by the direct binding of the stoichiometric inhibitors Mad2 and BubR1 to Cdc20, and their failure to bind APC^{Cdc20} strongly compromises spindle checkpoint function[5–7]. We therefore examined whether USP44-depleted cells are impaired in the formation of these APC-inhibitory complexes. To overcome the problem that USP44-depleted cells do not arrest in response to nocodazole, we used brief proteasome inhibition to arrest synchronized cells in mitosis (for detailed synchronization protocol see Supplementary Fig. 6a). Notably, in the absence of USP44, mitotic APC^{Cdc20} contained significantly reduced levels of co-purifying Mad2 (Fig. 3g and Supplementary Fig. 6b, c). As expected from the reduced Mad2–Cdc20 association, APC^{Cdc20} activity was increased in USP44-depleted cells despite checkpoint activation (Fig. 3f). These findings indicate that cells with reduced levels of USP44 fail to arrest in response to spindle damage because of their inability to maintain Mad2–Cdc20 complexes and consequent failure to restrain APC^{Cdc20} activity. Consistent with this hypothesis, the partial depletion of APC2 (anaphase-promoting complex subunit 2; also called ANAPC2) to levels that did not interfere with mitotic progression of otherwise unperturbed cells (Supplementary Fig. 7d) rescued the checkpoint defect of USP44-depleted cells (Fig. 3h and Supplementary Fig. 7). Together, our findings strongly suggest that USP44 maintains spindle checkpoint arrest by stabilizing the binding of Mad2 to APC^{Cdc20}.

USP44 antagonizes UbcH10-induced APC activation

USP44 might influence APC activity directly or indirectly. If direct, purified USP44 protein should affect APC activity *in vitro*. When

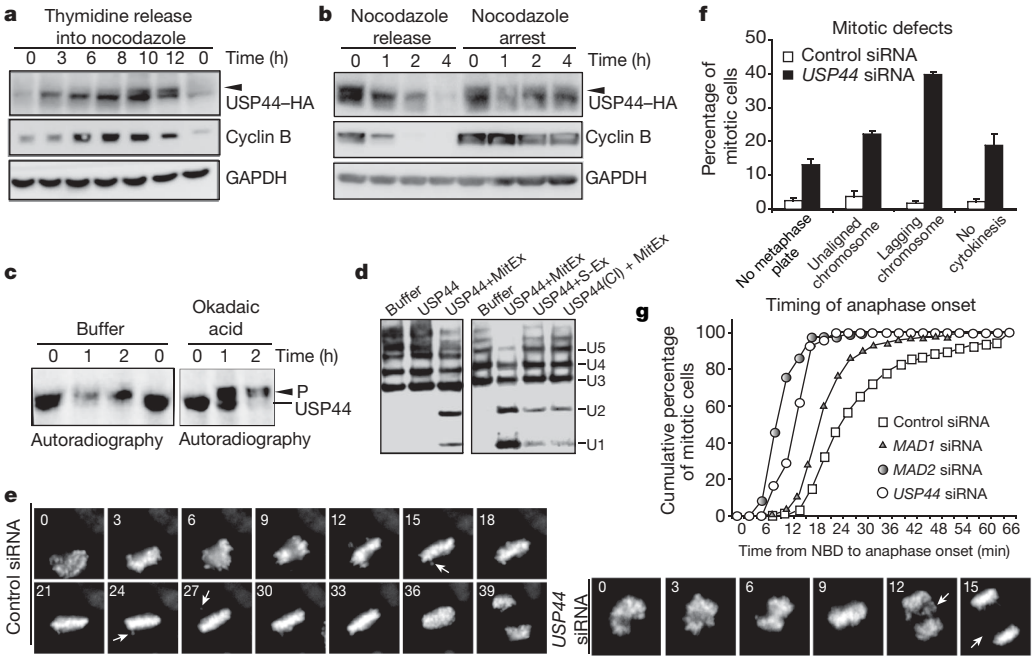


Figure 2 | USP44 activity is cell-cycle regulated and is required for proper spindle checkpoint function and anaphase timing. **a**, HeLa cells stably expressing USP44–HA were released from a double-thymidine block (0 h time point) into medium containing nocodazole. Cell lysates were blotted with the indicated antibodies. The arrowhead marks a slower migrating species of USP44 that is most prevalent in mitotic cells (10 h time point). **b**, HeLa cells stably expressing USP44–HA were synchronized in mitosis (100 ng ml^{-1} nocodazole), 10 mM cycloheximide was added (to prevent *de novo* protein synthesis), and cells were either released from (release) or cultured in the continued presence of nocodazole (arrest). Cell lysates were analysed by western blotting. **c**, Recombinant USP44 is phosphorylated in mitotic extracts. ^{35}S-labelled USP44 was incubated with mitotic extracts and treated either with buffer control or 25 nM okadaic acid (PP2A phosphatase inhibitor). **d**, Recombinant USP44 is activated by incubation with extracts from checkpoint-arrested HeLa cells. Flag-tagged USP44 and USP44(CI) (catalytic inactive, C282A) proteins were purified from insect cells. Where indicated, the bead-bound proteins were incubated for 3 h with mitotic extracts (MitEx) or S-phase extracts (S-Ex). *In vitro* deubiquitination assays using polyubiquitin chains (Ub_{3-7}) were performed for 20 min at 37 °C. The migration of monoubiquitin (U1), diubiquitin (U2) and longer ubiquitin chains is indicated. **e–g**, USP44 depletion leads to premature anaphase onset and segregation defects. HeLa cells stably expressing H2B–GFP were transfected with siRNAs and imaged by time-lapse microscopy 48 h after transfection. Times shown are in minutes. Arrows mark uncongressed chromosomes (control siRNA images) and lagging chromosomes (USP44 siRNA images). **f**, Mitotic defects were quantified ($n > 100$, error bars ± 1 s.d.). **g**, The kinetics of anaphase onset was determined by analysing the cumulative percentage of mitotic cells ($n > 100$). NBD, nuclear envelope breakdown.

activated USP44 was added to APC^{Cdc20} from checkpoint-arrested cells (CP-APC, which can be activated by the APC-specific E2 UbcH10[21,22]), there was strong inhibition of APC-dependent ubiquitination (Fig. 4a, b and Supplementary Fig. 8a). Notably, USP44 had no effect on ubiquitination reactions catalysed by APC^{Cdh1}, which is not subject to Mad2-mediated spindle checkpoint inhibition (Supplementary Fig. 8b). These findings show that USP44 can directly inhibit APC^{Cdc20} activity.

An accompanying study showed that APC-dependent multi-ubiquitination, but not proteolysis, leads to the disassembly of Mad2–Cdc20 complexes, and identified Cdc20 as a crucial ubiquitination target[16]. Notably, wild-type USP44 but not its catalytically inactive mutant was able to inhibit the UbcH10-driven dissociation of Mad2 from Cdc20 (Fig. 4b, lower panels), which prevented APC^{Cdc20} activation (Fig. 4b, upper panel) and CycB1 degradation (Fig. 4c, d). Moreover, USP44 was very potent at deubiquitinating Cdc20 *in vitro* (Fig. 5a, c, d) and the levels of multi-ubiquitinated Cdc20 increased after USP44 inhibition *in vivo* (Fig. 5e). Consistent with a function for Cdc20 multi-ubiquitination in Mad2–Cdc20 complex disassembly[16], the levels of Mad2 co-purifying with multi-ubiquitinated Cdc20 were substantially reduced in USP44-depleted cells (Fig. 5e). Whilst USP44 was much more efficient at deubiquitinating Cdc20 than cyclin B *in vitro* (Fig. 5a, b), we observed that higher levels of USP44 can promote weak deubiquitination of cyclin B (data not shown). Thus, although pointing towards Cdc20 being the critical target of USP44 that protects cells against premature APC-driven Mad2–Cdc20 complex disassembly, USP44 may additionally antagonize APC function by deubiquitinating other APC substrates.

Given that the APC-driven disassembly of Mad2–Cdc20 complexes requires the function of the E2 enzyme UbcH10 (ref. 16), an implication of our model is that reducing UbcH10 function should suppress the effects of low USP44 activity *in vivo*. In support of this hypothesis, depletion of UbcH10 in cells with reduced USP44 function decreased levels of Cdc20 multi-ubiquitination (Fig. 5e), restored Mad2–Cdc20 association (Fig. 5e), and consequently restored spindle checkpoint arrest (Fig. 5f and Supplementary Fig. 7). In conclusion, our *in vitro* and *in vivo* findings strongly suggest that USP44 directly antagonizes the APC-driven disassembly of Mad2–Cdc20 checkpoint complexes, probably by promoting the deubiquitination of Cdc20. Through the regulation of this ubiquitination–deubiquitination switch, USP44 prevents premature APC activation, which is crucial for proper mitotic timing and spindle checkpoint function.

Discussion

In this study we used a high-throughput shRNA-based forward genetic approach to screen for regulators of the spindle checkpoint and cell-cycle progression in human cells. Our screen identified the first deubiquitinating enzyme (USP44) controlling spindle checkpoint function and revealed a novel mode of APC regulation. USP44 prevents premature APC activation by inhibiting the disassembly of

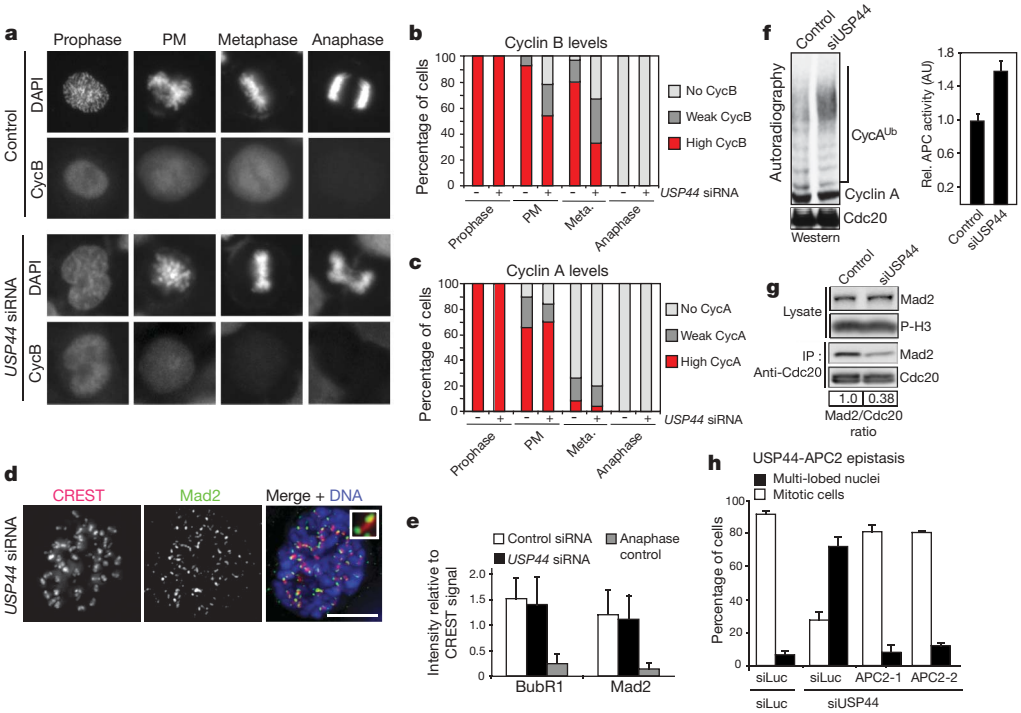


Figure 3 | USP44 regulates APC^{Cdc20} downstream of Mad2 recruitment to kinetochores. **a**–**c**, USP44-depleted cells prematurely degrade cyclin B but not cyclin A. HeLa cells were treated with siRNAs, synchronized and stained with the indicated antibodies for microscopic analysis. **a**, Representative images of cyclin B staining at different mitotic stages. DAPI, 4,6-diamidino-2-phenylindole; PM, prometaphase. **b**, **c**, Quantification of cyclin A and cyclin B levels ($n > 100$). **d**, USP44 depletion does not impair the localization of Mad2 to unattached kinetochores. Immunofluorescence images of prometaphase cells transfected with siRNAs against *USP44* were probed for Mad2 (green), CREST (red) and stained for DNA (blue). The inset shows the staining of one kinetochore pair at higher magnification. Scale bar, 10 μm. **e**, Quantification of Mad2 and BubR1 levels on prometaphase kinetochores ($n = 150$, error bars ±1 s.d.). **f**, USP44-depleted cells fail to restrain APC^{Cdc20} activation in response to spindle checkpoint activation. HeLa cells were treated with luciferase (control) or *USP44* siRNA, synchronized as described in Supplementary Fig. 6a, and mitotic cells collected by shake-off. Anti-Cdc27 immunoprecipitates (APC^{Cdc20}) were assayed for *in vitro* ubiquitination activity using ^{35}S-labelled cyclin A as a model substrate. The quantification of relative APC activity from three independent experiments is shown on the right (error bars indicate ±1 s.d.). **g**, Mad2–Cdc20 association is reduced after USP44 depletion. Cells were treated as described in **f**. Cdc20 immunoprecipitates and whole-cell lysates were blotted for the indicated proteins. **h**, Partial depletion of APC2 restores spindle checkpoint arrest to USP44-depleted cells. HeLa cells were synchronized and transfected with the indicated siRNAs as described in Supplementary Fig. 7a. Taxol (100 nM) was added after release from the second thymidine block and cells were fixed and stained for DNA (Hoechst) 16 h after release. The percentage of checkpoint-arrested (mitotic cells) and checkpoint-by-passing (multi-lobed nuclei) cells was quantified ($n > 100$, error bars ± s.d.).

Mad2–Cdc20 complexes, at least in part by antagonizing APC-dependent multi-ubiquitination of Cdc20 (Supplementary Fig. 9). These findings indicate that spindle checkpoint regulation of the APC is dynamic and based on a fine-tuned balance of counteracting ubiquitination and deubiquitination activities. On the basis of the critical role of USP44 in protecting cells from premature anaphase entry and to establish a common terminology across species, we propose to refer to deubiquitinating enzymes with similar APC antagonizing activities by the general name of 'protectins' (for sequence alignments with potential orthologues in other species, see Supplementary Figs 10 and 11).

In an accompanying study, it was found that the APC itself can drive the disassembly of checkpoint complexes[16]. This feed-forward activation loop provides an elegant mechanism to allow for the rapid activation of APC^{Cdc20} upon cessation of the checkpoint signal. However, it imposes the need for tight regulation before the completion of kinetochore attachment. We propose that protectins such as USP44 provide a critical safeguard mechanism to prevent the inappropriate activation of this feed-forward APC activation loop by antagonizing the multi-ubiquitination of Cdc20 and potentially other components of checkpoint-inhibited APC, thus maintaining the checkpoint-inhibited APC state (Supplementary Fig. 9). Although our data point towards CP-APC and especially Cdc20 as being critical targets of USP44, it is possible that USP44 more broadly antagonizes APC function by deubiquitinating other APC substrates such as cyclin B and securin.

An important question is how this ubiquitination–deubiquitination switch is regulated to promote the transition from the checkpoint-inhibited to the active APC state once each chromosome has achieved bipolar attachment. A potential clue comes from our finding that incubation of recombinant USP44 with extracts from checkpoint-arrested but not S-phase-arrested cells strongly increases its activity. How USP44 activity is increased in mitotic extracts is not yet clear. It

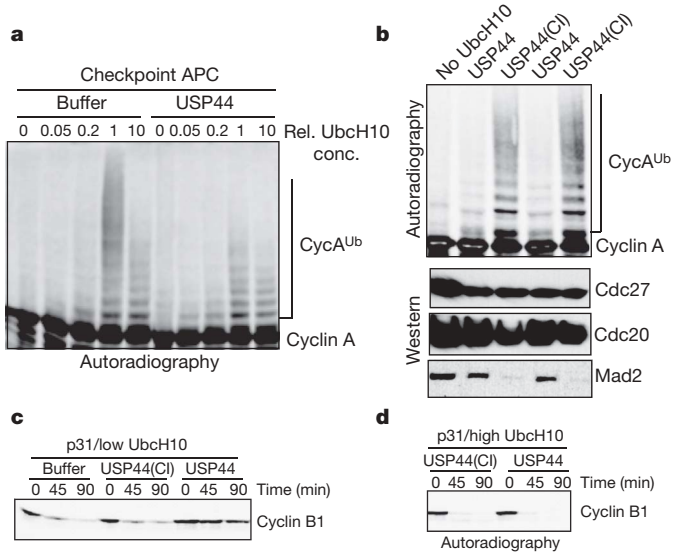


Figure 4 | USP44 inhibits APC^{Cdc20} activation *in vitro*. **a**, USP44 impairs the activation of checkpoint APC by UbcH10 *in vitro*. APC^{Cdc20} was immunopurified from checkpoint-arrested HeLa cells and assayed for *in vitro* ubiquitination activity using ^{35}S-labelled cyclin A as a substrate. UbcH10 and purified USP44 (after activation with mitotic extracts) were added where indicated. **b**, USP44 prevents UbcH10-induced Mad2–Cdc20 dissociation *in vitro*. APC^{Cdc20} was immunopurified from checkpoint-arrested HeLa cells and incubated with UbcH10 and p31 (which triggers Mad2–Cdc20 dissocation[16]) in combination with either wild-type or catalytically inactive USP44 (USP44(CI)). The amount of Mad2 bound to APC^{Cdc20} was determined by western blotting (lower panel), and APC^{Cdc20} activity (top panel) was analysed as described in Fig. 3f. **c**, **d**, USP44 and UbcH10 regulate CP-APC antagonistically *in vitro*. Extracts from checkpoint-arrested HeLa cells were incubated with the indicated proteins and the abundance of ^{35}S-labelled cyclin B was monitored over time by autoradiography. Note that whereas USP44 prevents CP-APC activation (as monitored by cyclin B degradation) induced by low levels of UbcH10 (**c**), fivefold higher levels of UbcH10 can overcome USP44's inhibitory effect (**d**), illustrating their antagonistic relationship.

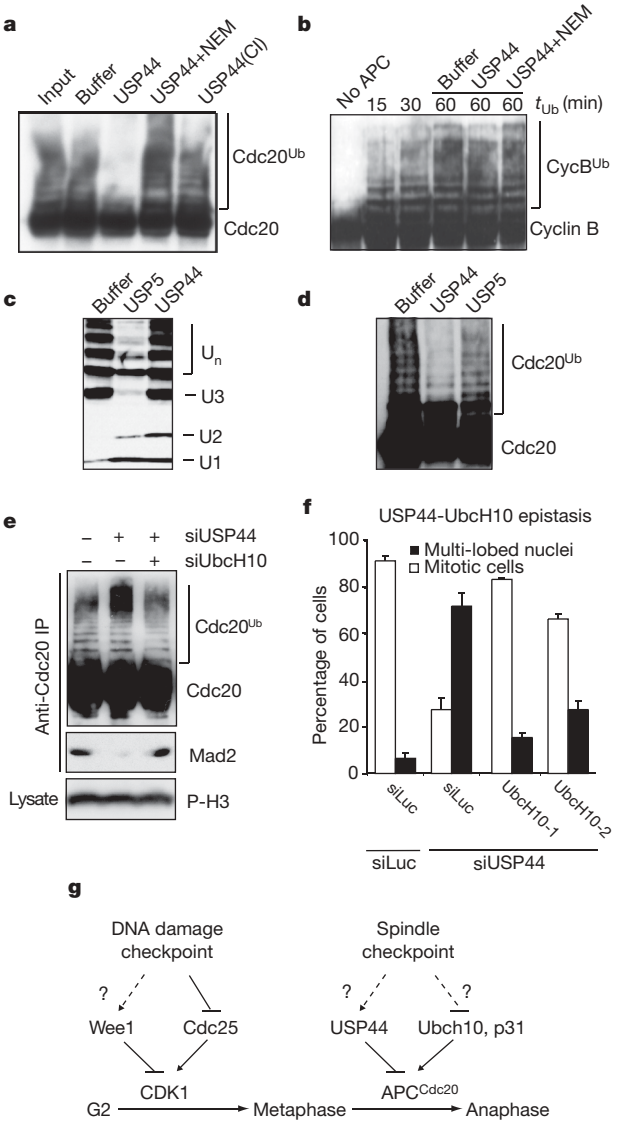


Figure 5 | USP44 deubiquitinates Cdc20 *in vitro* and *in vivo*. **a**, **b**, Ubiquitinated Cdc20 and cyclin B were immunopurified and incubated with USP44 (after activation with mitotic extract), USP44 inactivated by *N*-ethylmaleimide (NEM, a nonspecific deubiquitinating enzyme inhibitor), and USP44 with an active-site mutation (USP44(CI)) for 30 min. The reaction products were analysed by western blotting using anti-Cdc20 (**a**) or anti-cyclin B (**b**) antibodies. **c**, **d**, Comparison of *in vitro* deubiquitination activity of USP44 and USP5. *In vitro* deubiquitination assays using polyubiquitin chains (Ub_{3-7}) and multi-ubiquitinated Cdc20 were performed as described in Fig. 2d and panel **a**. Although USP5 exhibited high activity against free polyubiquitin chains (**c**), it failed to significantly deubiquitinate Cdc20 (**d**) at identical concentrations. **e**, USP44 and UbcH10 regulate Cdc20 multi-ubiquitination and Mad2–Cdc20 association antagonistically *in vivo*. Cells were treated with siRNAs, synchronized as described in Supplementary Fig. 6a, and only mitotic cells collected by shake-off. Cdc20 immunoprecipitates were blotted for the indicated proteins. **f**, Depletion of UbcH10 restores spindle checkpoint arrest to USP44-depleted cells. Cells were synchronized, transfected with siRNAs and treated as described in Supplementary Fig. 7a. The percentage of checkpoint-arrested (mitotic cells) and checkpoint-by-passing (multi-lobed nuclei) cells was quantified ($n > 100$, error bars ± 1 s.d.). **g**, Schematic model that illustrates the analogies between the regulation of entry into mitosis (G2–M transition) and the metaphase–anaphase transition.

could require the binding of a coactivator present only in mitotic cells. Alternatively, USP44 might be activated through phosphorylation by mitotic cyclin-dependent kinases or spindle checkpoint kinases. This would be an attractive possibility because it would allow for USP44 activity to be maximal during checkpoint arrest and then be rapidly inactivated once the checkpoint signal has been terminated. The inactivation of USP44, perhaps through dephosphorylation and subsequent degradation, coupled with a simultaneous increase in APC-mediated Cdc20 multi-ubiquitination promoted by UbcH10 and $p31^{comet}$, would trigger feed-forward activation of APC^{Cdc20} and rapid progression into anaphase (Supplementary Fig. 9).

A cell-cycle transition is a unidirectional change of state in which a cell that was performing one set of processes shifts its activity to perform a different set of processes. Inherent properties of these transitions are coordinated antagonistic circuits of positive and negative feedback loops that operate as a switch[23,24]. A well known example of this is the G2 to M transition in which the antagonistic functions of the activating Cdc25 phosphatase and the inhibitory Wee1 kinase regulate Cdk1 (ref. 25). Similarly, exit from mitosis in budding yeast is controlled by the opposing activities of the Cdk1 kinase and Cdc14 phosphatase[26]. No such switch mechanism has been described for the metaphase to anaphase transition, which is controlled by ubiquitination. On the basis of the results of our work and that of ref. 16, we propose that anaphase entry is controlled by a dynamic balance of ubiquitination and deubiquitination, which contributes to the switch-like transition from the checkpoint-inhibited to the active APC state. In this model, p31/UbcH10-stimulated ubiquitination functions analogously to dephosphorylation by Cdc25. Deubiquitination by protectins such as USP44 has a function that is analogous to phosphorylation by Wee1, and APC^{Cdc20} is analogous to Cdk1 (Fig. 5g). These studies illustrate the dynamic interaction between ubiquitination and deubiquitination in controlling key regulatory pathways, and should serve as a basis for future modelling of the metaphase to anaphase transition. We speculate that ubiquitination–deubiquitination switches similar to the one described here might contribute to the generation of bi-stable, switch-like transitions in other biological pathways.

METHODS

See Supplementary Information for full experimental details, including a detailed description of live cell time-lapse imaging and biochemical analyses.

shRNA screen and hit validation. HeLa cells were seeded in 96-well plates (2,000 cells per well) and co-transfected the following day with shRNA-expressing pSM2 plasmids (100–150 ng) and 10 ng dsRed-expressing vector using Exgen 500 (Fermentas). After 48 h, 100 nM Taxol was added and cells fixed with 4% formaldehyde 24 h after Taxol treatment. Cells were scored visually on an inverted microscope (Axiovert 200, Zeiss). The sequences of siRNAs used in this study are listed in Supplementary Table 3.

Cell synchronization. To isolate USP44-depleted mitotic cells, siRNA-transfected cells were synchronized at the G1/S transition by a double thymidine block. To this end, 4 h after siRNA transfection, cells were treated with thymidine (2.5 mM) for 16 h. After releasing cells for 7 h from the first thymidine block, thymidine was re-added for an additional 17 h. Cells were released from the second thymidine block into medium containing 100 ng ml^{-1} nocodazole to trigger spindle checkpoint activation. When cells started to enter mitosis (9 h after release), 10 μM MG132 was added to the medium for the last 2 h before collecting mitotic cells by shake-off (see schematic in Supplementary Fig. 6a).

Received 13 December 2006; accepted 19 February 2007.

1. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396,** 643–649 (1998).
2. Draviam, V. M., Xie, S. & Sorger, P. K. Chromosome segregation and genomic stability. *Curr. Opin. Genet. Dev.* **14,** 120–125 (2004).
3. Kops, G. J., Weaver, B. A. & Cleveland, D. W. On the road to cancer: aneuploidy and the mitotic checkpoint. *Nature Rev. Cancer* **5,** 773–785 (2005).
4. Peters, J. M. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nature Rev. Mol. Cell Biol.* **7,** 644–656 (2006).
5. Musacchio, A. & Hardwick, K. G. The spindle checkpoint: structural insights into dynamic signalling. *Nature Rev. Mol. Cell Biol.* **3,** 731–741 (2002).
6. Bharadwaj, R. & Yu, H. The spindle checkpoint, aneuploidy, and cancer. *Oncogene* **23,** 2016–2027 (2004).
7. Nasmyth, K. How do so few control so many? *Cell* **120,** 739–746 (2005).
8. Hoyt, M. A., Totis, L. & Roberts, B. T. S. *cerevisiae* genes required for cell cycle arrest in response to loss of microtubule function. *Cell* **66,** 507–517 (1991).
9. Li, R. & Murray, A. W. Feedback control of mitosis in budding yeast. *Cell* **66,** 519–531 (1991).
10. Weiss, E. & Winey, M. The *Saccharomyces cerevisiae* spindle pole body duplication gene MPS1 is part of a mitotic checkpoint. *J. Cell Biol.* **132,** 111–123 (1996).
11. Karess, R. Rod-Zw10-Zwilch: a key player in the spindle checkpoint. *Trends Cell Biol.* **15,** 386–392 (2005).
12. Habu, T., Kim, S. H., Weinstein, J. & Matsumoto, T. Identification of a MAD2-binding protein, CMT2, and its role in mitosis. *EMBO J.* **21,** 6419–6428 (2002).
13. Xia, G. *et al.* Conformation-specific binding of p31(comet) antagonizes the function of Mad2 in the spindle checkpoint. *EMBO J.* **23,** 3133–3143 (2004).
14. Mapelli, M. *et al.* Determinants of conformational dimerization of Mad2 and its inhibition by p31comet. *EMBO J.* **25,** 1273–1284 (2006).
15. Kops, G. J. *et al.* ZW10 links mitotic checkpoint signaling to the structural kinetochore. *J. Cell Biol.* **169,** 49–60 (2005).
16. Reddy, S. K., Rape, M., Marganski, W. A. & Kirschner, M. W. Ubiquitination by the anaphase-promoting complex drives spindle checkpoint inactivation. *Nature* advance online publication, doi:10.1038/nature05734 (this issue).
17. Silva, J. M. *et al.* Second-generation shRNA libraries covering the mouse and human genomes. *Nature Genet.* **37,** 1281–1288 (2005).
18. Meraldi, P., Draviam, V. M. & Sorger, P. K. Timing and checkpoints in the regulation of mitotic progression. *Dev. Cell* **7,** 45–60 (2004).
19. Geley, S. *et al.* Anaphase-promoting complex/cyclosome-dependent proteolysis of human cyclin A starts at the beginning of mitosis and is not subject to the spindle assembly checkpoint. *J. Cell Biol.* **153,** 137–148 (2001).
20. den Elzen, N. & Pines, J. Cyclin A is destroyed in prometaphase and can delay chromosome alignment and anaphase. *J. Cell Biol.* **153,** 121–136 (2001).
21. Rape, M. & Kirschner, M. W. Autonomous regulation of the anaphase-promoting complex couples mitosis to S-phase entry. *Nature* **432,** 588–595 (2004).
22. Rape, M., Reddy, S. K. & Kirschner, M. W. The processivity of multiubiquitination by the APC determines the order of substrate degradation. *Cell* **124,** 89–103 (2006).
23. Ferrell, J. E. & Xiong, W. Bistability in cell signaling: How to make continuous processes discontinuous, and reversible processes irreversible. *Chaos* **11,** 227–236 (2001).
24. Markevich, N. I., Hoek, J. B. & Kholodenko, B. N. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J. Cell Biol.* **164,** 353–359 (2004).
25. Takizawa, C. G. & Morgan, D. O. Control of mitosis by changes in the subcellular location of cyclin-B1-Cdk1 and Cdc25C. *Curr. Opin. Cell Biol.* **12,** 658–665 (2000).
26. Stegmeier, F. & Amon, A. Closing mitosis: the functions of the Cdc14 phosphatase and its regulation. *Annu. Rev. Genet.* **38,** 203–232 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Taylor, H. Yu, W. Earnshaw and J. Jin for gifts of reagents; M. Vidal for providing access to their BioRobot platform; S. Lyman and R. King for communicating unpublished results and assistance with the development of the Taxol screening assay; C. Shamu for access to the ICCB-Longwood screening facilities; S. Reddy for helpful comments throughout the course of the work; and T. Westbrook and A. Smogorzewska for their critical reading of the manuscript. F.S. is a fellow of the Helen Hay Whitney Foundation. M.R. is a Human Frontiers Science Program Long-Term Fellow. The siRNA and ICCB-Longwood resources used were funded in part by a NCI grant (T. Mitchison). M.E.S. is an American Cancer Society Postdoctoral Fellow. X.L.A. is an NIH pre-doctoral fellow. M.W.K. thanks the National Institute of General Medical Sciences for its support for the grant Cell Cycle Regulation. This work was supported by grants from NIH and DOD to S.J.E. and by grants from the NIH to J.W.H. S.J.E. is an investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.J.E (selledge@genetics.med.harvard.edu), J.W.H (wade_harper@hms.harvard.edu), or M.W.K. (marc@hms.harvard.edu).

LETTERS

Giant cladoxylopsid trees resolve the enigma of the Earth's earliest forest stumps at Gilboa

William E. Stein[1], Frank Mannolini[2], Linda VanAller Hernick[2], Ed Landing[2] & Christopher M. Berry[3]

The evolution of trees of modern size growing together in forests fundamentally changed terrestrial ecosystems[1–3]. The oldest trees are often thought to be of latest Devonian age (about 380–360 Myr old) as indicated by the widespread occurrence of *Archaeopteris* (Progymnospermopsida)[4]. Late Middle Devonian fossil tree stumps, rooted and still in life position, discovered in the 1870s from Gilboa, New York[5], and later named *Eospermatopteris*, are widely cited as evidence of the Earth's 'oldest forest'[6,7]. However, their affinities and significance have proved to be elusive because the aerial portion of the plant has been unknown until now. Here we report spectacular specimens from Schoharie County, New York, showing an intact crown belonging to the cladoxylopsid *Wattieza* (Pseudosporochnales)[8] and its attachment to *Eospermatopteris* trunk and base. This evidence allows the reconstruction of a tall (at least 8 m), tree-fern-like plant with a trunk bearing large branches in longitudinal ranks. The branches were probably abscised as frond-like modules. Lower portions of the trunk show longitudinal carbonaceous strands typical of *Eospermatopteris*, and a flat bottom with many small anchoring roots. These specimens provide new insight into Earth's earliest trees and forest ecosystems. The tree-fern-like morphology described here is the oldest example so far of an evolutionarily recurrent arborescent body plan within vascular plants. Given their modular construction, these plants probably produced abundant litter, indicating the potential for significant terrestrial carbon accumulation and a detritus-based arthropod fauna by the Middle Devonian period.

Although the fossil record provides evidence of increasing size and complexity of plants through the Devonian, little is known about how the origin of tree-sized individuals changed the evolutionary dynamics of terrestrial ecosystems. The famous 'earliest forest' occurrence of upright and rooted sandstone casts called *Eospermatopteris* found at three stratigraphic levels (Middle Devonian, upper Givetian to about the Middle/Upper Devonian boundary, about 385 Myr ago) at Gilboa, New York, displays typical difficulties encountered in making morphological and ecological interpretations. Each cast, invariably broken 50–150 cm above the base, is typically 50–100 cm in circumference, flaring proximally to as much as 330 cm at the base. A distinct pattern of longitudinally oriented carbonaceous strands occurs on the trunk, often showing proximal anastomosis. *Eospermatopteris* was initially thought to be associated with isolated branch systems and was reconstructed as a pteridosperm (early seed plant)[9]. However, subsequent work has provided little support for this reconstruction[10]. Thus, evidence so far allows wide latitude in phylogenetic and ecological interpretations of the plant, including aneurophytalean progymnosperm[11], lepidosigillarioid lycopsid[1,12] or cladoxylopsid[13].

Among these possibilities, it is becoming increasingly evident that members of the class Cladoxylopsida, ranging at least from the early Middle Devonian (Eifelian) into the Carboniferous, were major contributors to floras worldwide[14]. Traditionally considered intermediate between Lower Devonian vascular plants and ferns or sphenopsids, we do not yet understand these plants well enough to reconstruct their overall architecture or phylogeny accurately. Recent work indicates that the group may have had considerable morphological and anatomical diversity. Several forms were probably quite large and some had significant secondary tissues[15–19]. Within Cladoxylopsida, the order Pseudosporochnales, represented especially by the genus *Pseudosporochnus* from Goé, Belgium (latest Eifelian), are among the best-studied[20,21]. Compressions are commonly found of digitately branched systems bearing sterile and fertile non-laminar appendages, all having a highly distinctive 'speckled' pattern caused by sclereid nests in the cortex. On the basis of the association of typical branches with trunk fragments, *Pseudosporochnus* has been reconstructed as a small tree[22]. However, only one specimen shows direct insertion of a single fragmentary branch base on the trunk. Well-defined surfaces at the proximal end of branches clearly match the regular pattern of attachment scars on the trunks. From this, regular abscission of the branches from the trunk has been inferred[22]. The genus *Wattieza* from Belgium and Venezuela shows a similar construction of branch systems, including probable basal abscission surfaces and speckled texture, but bears more complex appendages, some with recurved tips bearing sporangia[8]. However, the trunk remains unknown.

The material described here is derived from a New York State Department of Environmental Conservation quarry on the northwest slope of South Mountain, Schoharie County, New York (42° 23′ N, 74° 16′ W). This site is located about 13 km east of the highest *in situ* Gilboa stump horizon at Manorkill Falls[7] and lies close to inferred palaeo-shoreline[23]. Long-term collecting at this locality has produced a diverse flora[24].

The quarry exposes 7.5 m of rock in the Oneonta Formation, the lowest formation of the eastern terrestrial facies of the Genesee Group. Although the dating of the terrestrial sediments remains imprecise, palynostratigraphy of nearby samples yields latest Givetian to earliest Frasnian age[25]. The lower 3.3-m exposure includes terrestrial red/brown siltshales, sandstones with root structures, and several soil horizons. The highest soil horizon is overlain by a sandstone unit with west-dipping foresets (3.3–6.0 m), which is consistent with a small fluvial delta. Abundant plant debris occurs low in these foresets, including shoot systems and logs ranging from a few centimetres to several metres in length. The logs mostly lie in near north–south orientation perpendicular to the palaeocurrent direction and parallel with small wave ripples. Some specimens show attachment of fine sterile or fertile appendages to branches. These data indicate minimal transport and rapid deposition. We describe two exceptionally complete specimens here.

[1]Department of Biological Sciences, Binghamton University, Binghamton, New York 13902-6000, USA. [2]New York State Museum, Albany, New York 12230, USA. [3]School of Earth, Ocean and Planetary Sciences, Cardiff University, Cardiff CF10 3YE, UK.

The first (Fig. 1a, b) represents the only tree-sized member of the Pseudosporochnales so far discovered with substantial connection of the aerial portions of the plant. More than 2 m of trunk was found in the field, of which the upper 130 cm has been recovered mostly intact. The trunk ranges in width from 15 cm at the top, within a well-defined crown region, to 18 cm at the recovered specimen's proximal end. At least eight digitate branches are attached in the crown, but several more were probably attached both above and below the exposed surface of the compression. The branches bear sterile and fertile appendages (Fig. 1c). The appendages have a central axis and whorled dichotomous ultimate units. Fertile appendages bear recurved tips bearing sporangia. These features allow identification to *Wattieza* Stockmans[8]. Attached branches observed in face-on and edge-on views seem to have been bilaterally symmetrical, giving the plant an overall appearance similar to that of a modern tree-fern with large fronds. Within the zone of obvious attachment, the branches occur in orthostichous ranks, and a regular pattern of branch scars continues down the entire preserved length of trunk (Fig. 1b). Just below the crown, isolated branches occur in the matrix along the right side of the trunk aligned as if still attached. This suggests an extensive, mostly ascendant, set of branches borne by the plant at the time of death. Further down, the trunk shows no evidence of attached senescent branches, and branch scars become progressively less distinct. The latter suggests the remodelling of outer tissues of the trunk after branch fall. The trunk was found within and slightly above bedding planes showing detached and randomly arranged branches. Detached branches match attached branches in morphology and size and thus potentially comprise litter derived from this individual. Some detached branches show likely abscission surfaces generally comparable to those described previously in *Pseudosporochnus* and *Wattieza*[22].

The second specimen, a large trunk (Fig. 2a), was found about 2 m away and seems to be mostly proximal to the first specimen, but with some overlap. This specimen is a flattened internal cast with both upper and lower counterparts showing surface features. The trunk is more than 6 m in length, with a width of 13 cm at the top and expanding to 47 cm near the base. The top is truncated but shows an identical pattern of branch scars and speckled surface to that observed about 75 cm below the crown in the first specimen (Fig. 2a, 3a). Below this level, the trunk shows diminution of branch attachment features, as in the first specimen, replaced by a system of longitudinal coalified strands. The coalified strands are distally narrow and closely spaced, becoming larger and more separate towards the base. Near its proximal end, about 70 cm from the base of the specimen (Fig. 2a, 3b), the trunk swells to form a flattened cast (8–10 cm thick), with a change in the pattern of the longitudinal strands indicating anastomosis. Combined evidence leaves little doubt about assignment of the specimen to *Eospermatopteris*. Unlike the upright *Eospermatopteris* stumps at Gilboa, however, this trunk is preserved along a bedding plane and is radially compressed. We think this difference very probably explains the absence of a bulbous base in this specimen as opposed to that observed in some, but not all, Gilboa stumps. Below the interpreted flat bottom of the trunk, many coalified strands extend somewhat sinuously into the matrix, suggesting attached roots.

The union of *Wattieza* branches with a trunk and base identical to those of *Eospermatopteris* provides the first direct evidence of general body form including attachment for both genera, and allows new insight into the enigmatic Gilboa trees. This evidence also provides the first reasonably complete picture of the size and architecture of large pseudosporochnalean cladoxylopsids (Fig. 2b). Evidence from the two specimens combined indicates that the height of the tree may have exceeded 8 m. The basal diameter of our specimen falls within the range of stump sizes observed at Gilboa. However, the largest stumps at Gilboa are twice the diameter of ours, suggesting even greater height.

Cladoxylopsids have long been interpreted as intermediates between the earliest vascular plants and living pteridophytes. It is

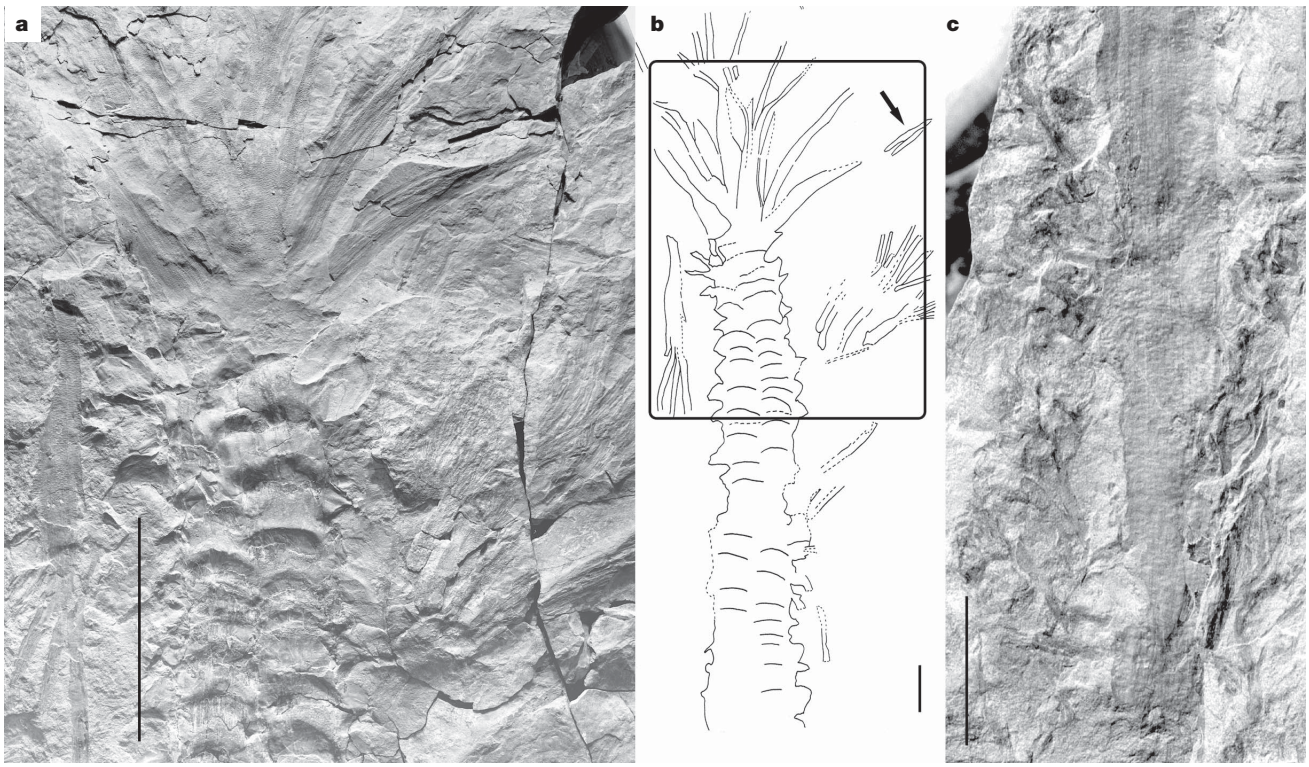


Figure 1 | *Wattieza* Stockmans from South Mountain, New York, NYSM 17039. a, General view of the crown portion, showing longitudinal ranks of branch bases on the trunk proximally, and attached branches with digitate ramification and speckled surface pattern distally. Scale bar, 20 cm. **b**, Line drawing of the specimen as recovered including trunk and crown; the box shows the portion in **a**, and the arrow indicates the branch in **c**. Scale bar, 10 cm. **c**, Close-up of a distal branch showing speckled texture and lateral appendages. Scale bar, 20 mm.

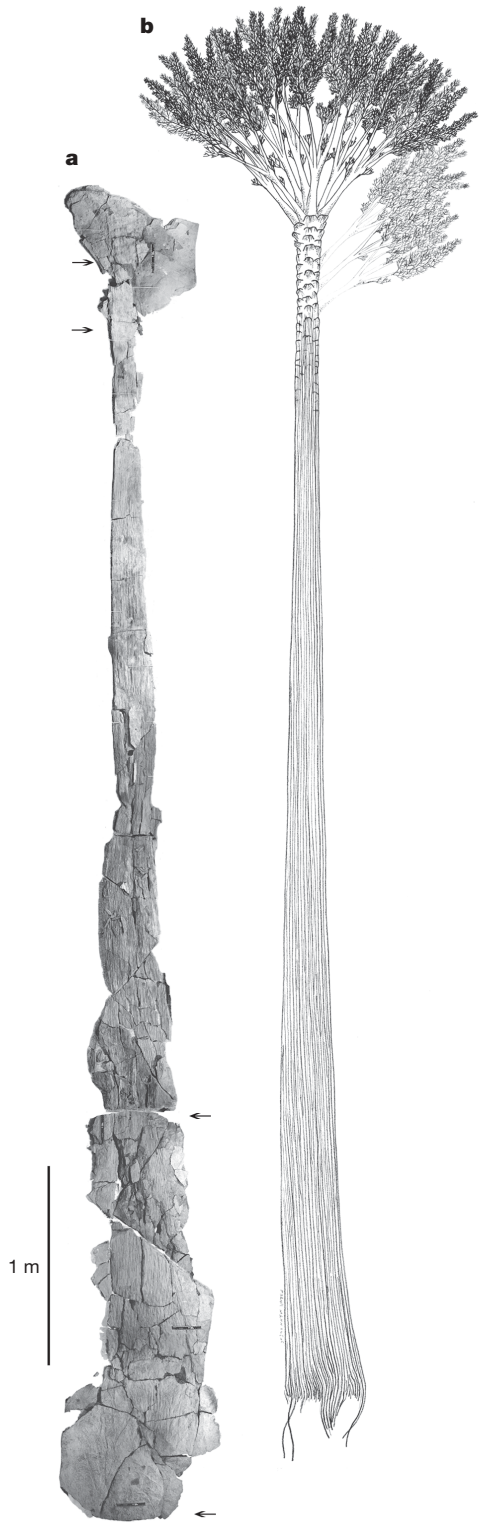


Figure 2 | Large trunk and reconstruction of plant from South Mountain, New York. a, Composite image of large trunk specimen, a cast with upper and lower counterparts, NYSM 17040. Arrows at the distal end (top) correspond to the region in Fig. 3a; arrows at the proximal end (bottom) correspond to the region in Fig. 3b. **b**, Line drawing showing the architecture of *Wattieza* attached to *Eospermatopteris*. The length of the trunk is not firmly established, so the minimum tree height is shown. Light branches right, also in Fig. 1a right, appear in life position but are not definitively attached. Scale bar, 1 m for both panels.

therefore interesting to see how instantly recognizable and, in a significant sense, 'modern' the tree-like architecture of *Wattieza* seems to be. Indeed, phylogenetically divergent modern forms including tree-ferns, cycads and palms are fundamentally similar in structure. This body plan now stands unequivocally as the oldest known arborescent terrestrial plant form, with trunks of similar diameter (at least 13 cm) showing attachment scars known from the lower Middle Devonian (Goé, Eifelian)[22]. The *Wattieza* body plan contrasts sharply with more diffusely branched *Archaeopteris*, conifers and most dicotyledonous angiosperm trees.

In addition, it is evident that this early tree architecture is not merely primitive but instead has evolved recurrently within vascular plants. It is perhaps significant that arborescence in *Wattieza* is associated with trunk bases lacking taproots, seeming instead to be anchored by many roots of nearly equal size. We recognize this syndrome in the modern analogues, where an extensive root mantle in some contributes significantly to their self-supporting habit. One previously described cladoxylopsid is known to exhibit a root mantle, although in this instance the plant is apparently not arborescent[26]. It remains to be determined whether *Eospermatopteris* also has a root mantle at the base. Perhaps this morphology represents a wetland ecological specialization distinguishing pseudosporochnalean cladoxylopsids of Gilboa from riparian conifer-like *Archaeopteris*[7]. So far, however, neither kind of root has been adequately circumscribed as a morphological, ecological or biogeochemical entity. Caution is

Figure 3 | Trunk top and base of *Wattieza* Stockmans from South Mountain, New York; details of the large trunk specimen, NYSM 17040. a, Counterpart of the distal portion of the trunk indicated by arrows in Fig. 2a, having speckled texture and prominent branch scars. Scale bar, 10 cm. **b**, Lower counterpart of the proximal cast region of trunk indicated by arrows in Fig. 2a, with the cast removed. The specimen shows longitudinal carbonaceous strands that anastomose at the base of the trunk (arrow); below, roots appear converging and loose in the matrix. Scale bar, 50 cm.

therefore advised in using the above general ecological characterizations for estimating weathering or carbon cycling in the Devonian.

Middle Devonian forests including pseudosporochnalean cladoxylopsid trees were probably widespread in a low south latitude warm-temperate zone along the Laurentian and northern European margins of the Acadian mountain belt, as well as in adjacent portions of Gondwana including South America[27]. Living environments for pseudosporochnaleans represented by the Gilboa trees may have been waterlogged and poorly oxygenated[7]. However, evidence for this conclusion must necessarily come from detailed sedimentological analysis, not from gross morphology of the root system or the identity of the plants at high taxonomic level. Given varied ecology in modern analogues, and spatial arrangement of *in situ* stumps[7], pseudosporochnalean trees may have formed a partial or complete canopy in some settings. However, because these plants lacked laminar leaves, there may have been less shading than is observed in modern tree-fern forests. The mode of growth of these trees, involving the shedding of large photosynthetic–reproductive modules, strongly implies the abundant production of litter. This would permit significant carbon accumulation with possible global effects on carbon budgets and atmospheric pCO_2. Branch litter of this magnitude, with a mixed texture comprised of both coarser and finer elements, would seemingly encourage the development of a detritus-based arthropod fauna.

Received 13 November 2006; accepted 26 February 2007.

1. Algeo, T. J., Scheckler, S. E. & Maynard, J. B. in *Plants Invade the Land: Evolutionary and Environmental Approaches* (eds Gensel, P. G. & Edwards, D.) 213–236 (Columbia Univ. Press, New York, 2001).
2. Driese, S. G. & Mora, C. I. in *Plants Invade the Land: Evolutionary and Environmental Approaches* (eds Gensel, P. G. & Edwards, D.) 237–253 (Columbia Univ. Press, New York, 2001).
3. Greb, S. F., DiMichele, W. A. & Gastaldo, R. A. in *Wetlands Through Time* (eds DiMichele, W.A., & Greb, S.) 1–40 (Geol. Soc. Am. Special Paper no. 399, Boulder, Colorado, 2006).
4. Meyer-Berthaud, B., Scheckler, S. E. & Wendt, J. *Archaeopteris* is the earliest known modern tree. *Nature* **938,** 700–701 (1999).
5. Dawson, J. W. On new tree ferns and other fossils from the Devonian. *Q. J. Geol. Soc. Lond.* **27,** 269–275 (1871).
6. Goldring, W. The oldest known petrified forest. *Sci. Mthly* **24,** 514–529 (1927).
7. Driese, S., Mora, C. I. & Elick, J. M. Morphology and taphonomy of root and stump casts of the earliest trees (Middle to Late Devonian), Pennsylvania and New York, USA. *Palaios* **12,** 524–537 (1997).
8. Berry, C. M. A reconsideration of *Wattieza* Stockmans (here attributed to Cladoxylopsida) based on a new species from the Devonian of Venezuela. *Rev. Palaeobot. Palynol.* **112,** 125–146 (2000).
9. Goldring, W. The Upper Devonian forest of seed ferns. *NY State Mus. Bull.* **521,** 50–72 (1924).
10. Kräusel, R. & Weyland, H. Pflanzenreste aus dem Devon IX. Ein Stamm von *Eospermatopteris* Bau aus dem Mittledevon des Kirberges, Elberfeld. *Senckenbergiana* **17,** 1–20 (1935).
11. Beck, C. B. Current status of the progymnospermopsida. *Rev. Palaeobot. Palynol.* **21,** 5–23 (1976).
12. Pigg, K. B. Evolution of Isoetalean lycopsids. *Ann. Mo. Bot. Gdn* **79,** 589–612 (1992).
13. Boyer, J. S. & Matten, L. C. Anatomy of *Eospermatopteris eriana* from the Upper Middle Devonian (=Givetian) of New York. *Int. Org. Palaeobot.* **5,** 13 (1996).
14. Berry, C. M. & Fairon-Demaret, M. in *Plants Invade the Land: Evolutionary and Environmental Approaches* (eds Gensel, P. G. & Edwards, D.) 120–158 (Columbia Univ. Press, New York, 2001).
15. Lemoigne, Y. & Iurina, Y. A. *Xenocladia medullosina* Ch. A. Arnold (1940) 1952 du Dévonien moyen du Kazakhstan (URSS). *Geobios* **16,** 513–547 (1983).
16. Stein, W. E. & Hueber, F. M. The anatomy of *Pseudosporochnus: P. hueberi* from the Devonian of New York. *Rev. Palaeobot. Palynol.* **60,** 311–359 (1989).
17. Berry, C. M. & Yi, W. A new plant attributed to Cladoxylopsida from the Middle Devonian of Yunnan province, China. *Rev. Palaeobot. Palynol.* **142,** 63–78 (2006).
18. Soria, A., Meyer-Berthaud, B. & Scheckler, S. E. Reconstructing the architecture and growth habit of *Pietschia levis* sp. nov. (Cladoxylopsida) from the Late Devonian of southeastern Morocco. *Int. J. Plant Sci.* **162,** 911–926 (2001).
19. Hilton, J., Geng, B. & Kenrick, P. A novel late Devonian (Frasnian) woody cladoxylopsid from China. *Int. J. Plant Sci.* **164,** 793–805 (2003).
20. Leclercq, S. & Banks, H. P. *Pseudosporochnus nodosus* sp. nov., a Middle Devonian plant with cladoxylalean affinities. *Palaeontographica B* **110,** 1–34 (1962).
21. Berry, C. M. & Fairon-Demaret, M. A reinvestigation of the cladoxylopsid *Pseudosporochnus nodosus* Leclercq et Banks from the Middle Devonian of Goé, Belgium. *Int. J. Plant Sci.* **158,** 350–372 (1997).
22. Berry, C. M. & Fairon-Demaret, M. The architecture of *Pseudosporochnus nodosus* Leclercq et Banks: a Middle Devonian cladoxylopsid from Belgium. *Int. J. Plant Sci.* **163,** 699–713 (2002).
23. Bridge, J. S. & Willis, B. J. Marine transgressions and regressions recorded in Middle Devonian shore-zone deposits of the Catskill clastic wedge. *Geol. Soc. Am. Bull.* **106,** 1440–1458 (1994).
24. Banks, H. P., Grierson, J. D. & Bonamo, P. M. in *The Catskill Delta* (eds Woodrow, D. L. & Sevon, W. D.) 125–141 (Geol. Soc. Am. Special Paper no. 201, Boulder, Colorado, 1985).
25. Traverse, A. & Schuyler, A. Palynostratigraphy of the Catskill and part of the Chemung magnifacies, southern New York State, USA. *Courier Forschungsinstitut Senckenberg* **169,** 261–274 (1994).
26. Soria, A. & Meyer-Berthaud, B. Tree fern growth strategy in the late Devonian cladoxylopsid species *Pietzschia levis* from the study of its stem and root system. *Am. J. Bot.* **91,** 10–23 (2004).
27. Hammond, S. E. & Berry, C. M. A new species of *Tetraxylopteris* (Aneurophytales) from the Devonian of Venezuela. *Bot. J. Linn. Soc.* **148,** 275–303 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support was provided by the New York State Museum.

Author Contributions L.V.H. and F.M. were responsible for initial discovery, field work and museum curation, E.L. for geological interpretation, W.E.S. and C.M.B. for palaeobotanical interpretation, F.M. for drawing the reconstruction in Fig. 2, and W.E.S. for writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.E.S. (stein@binghamton.edu) or C.M.B. (chris.berry@earth.cf.ac.uk).

LETTERS

A stepwise mechanism for acetylcholine receptor channel gating

Prasad Purohit[1], Ananya Mitra[1] & Anthony Auerbach[1]

Muscle contraction is triggered by the opening of acetylcholine receptors at the vertebrate nerve–muscle synapse[1–4]. The M2 helix of this allosteric membrane protein lines the channel, and contains a 'gate' that regulates the flow of ions through the pore. We used single-molecule kinetic analysis to probe the transition state of the gating conformational change and estimate the relative timing of M2 motions in the α-subunit of the murine acetylcholine receptor[5]. This analysis produces a 'Φ-value' for a given residue that reflects its open-like versus closed-like character at the transition state. Here we show that most of the residues throughout the length of M2 have a Φ-value of ~0.64 but that some near the middle have lower Φ-values of 0.52 or 0.31, suggesting that αM2 moves in three discrete steps. The core of the channel serves both as a gate that regulates ion flow and as a hub that directs the propagation of the gating isomerization through the membrane domain of the acetylcholine receptor.

Nicotinic acetylcholine receptors (AChRs) are ion channels that gate between ion-impermeable closed (C) and ion-permeable open (O) conformations. The membrane domain of each of the five AChR subunits has four helices (M1–M4). The structure of the unliganded-closed *Torpedo* AChR (at 4 Å resolution) shows that a narrow part of the channel is formed by residues at the mid-section of the second transmembrane helix, M2[6,7]. Here, interactions between side chains in different subunits form a 'hydrophobic girdle' that presumably shuts off ion flow in the C conformation and, hence, might act as a gate[6]. Results obtained with substituted cysteine[8,9] or histidine[10] accessibility methods suggest that the main barrier to the diffusion of ions in closed AChRs is near the cytoplasmic limit of M2. However, similar experiments with other Cys-loop receptors indicate that in these closely-related proteins it is the middle of M2 that serves this function[11,12]. Computational studies[13] support the idea that the middle of M2 is a barrier to ionic conduction, and mutagenesis[14] and accessibility[15] studies show that M2 equatorial residues move during C↔O gating, as required for a gate.

The αM2 helix comprises 27 residues that run from the intracellular limit of the membrane (1′) to the extracellular domain (27′; Fig. 1). The character of the di-liganded, C↔O transition state ensemble (TSE) was probed by measuring the single-channel opening (k_o) and closing (k_c) rate constants for sets of side chain substitutions of individual amino acids. The intermediates of the gating reaction are too brief to be observed directly but some properties of the TSE can be deduced from Φ, the slope of a log–log plot of k_o versus K_{eq} (k_o/k_c) for each mutational series[16]. One interpretation of Φ is that it gives the relative timing of the movement of the perturbed residue, with higher values indicating earlier, and similar values indicating temporally-correlated, gating motions[5] (see Methods).

Φ-values have been reported previously for three αM2 residues, S27′ (0.64), V17′ (0.61) and L9′ (0.26)[17]. The step decrease in Φ in αM2 (0.35 units between 17′ and 9′) is roughly similar in magnitude and location to that in δM2 (0.25 units between 12′ and 9′)[18]. We interpret these breakpoints in Φ to indicate that in the α- and δ-subunits the upper part of M2 moves before the middle. That is, in both subunits the mid-section of M2 appears to be a flexible junction.

We focused our attention on the αM2 residues between 16′ and 2′, a region that forms a narrow bore of the channel and brackets the presumptive gate[6]. Figure 2 shows Φ-value analyses for 12 αM2 positions (Supplementary Fig. 1 and Supplementary Tables 1 and 2). Mutations of three additional positions (S10′, I5′ and T2′) exhibited only wild-type gating kinetics. This result makes it unlikely that these residues, one of which sets the conductance of the open channel (2′)[19], are moving parts of a 'gate', but does not rule out gating motions below 2′[9,10,20]. We cannot conclude definitively that a small excursion in K_{eq} indicates that there is relatively less gating motion (for example, at the cytoplasmic end of M2), because little or no change may be apparent if a side chain and its local environment move in register, and because the magnitude of the change in K_{eq} may not reflect the magnitude of a residue's motion.

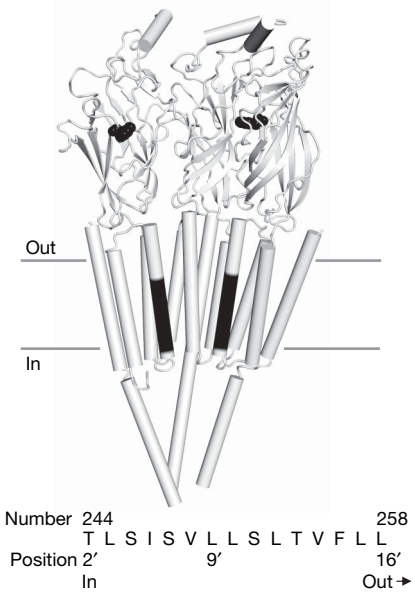


Figure 1 | The location of αM2. The α_δ-, ε- and α_ε-subunits of the AChR; for clarity the δ- and β-subunits are not shown. The thick horizontal lines mark the approximate location of the membrane. The M2 helix forms the lumen of the channel. In each α-subunit, in black is residue W149 at the two transmitter binding sites and residues 17′–2′ (top to bottom) in each αM2 helix. The distance from W149 to the 9′ residue of αM2 is ~60 Å (*Torpedo* AChR, PDB database accession number 2bg9.pdb). Below, the amino acid sequence (N-to-C) of M2 in the mouse α-subunit. K0′ and E−1′ are likely to be in the cytoplasmic domain and E20′ is likely to be the first residue in the extracellular ionic compartment.

[1]Department of Physiology and Biophysics, State University of New York at Buffalo, Buffalo, New York 14214, USA.

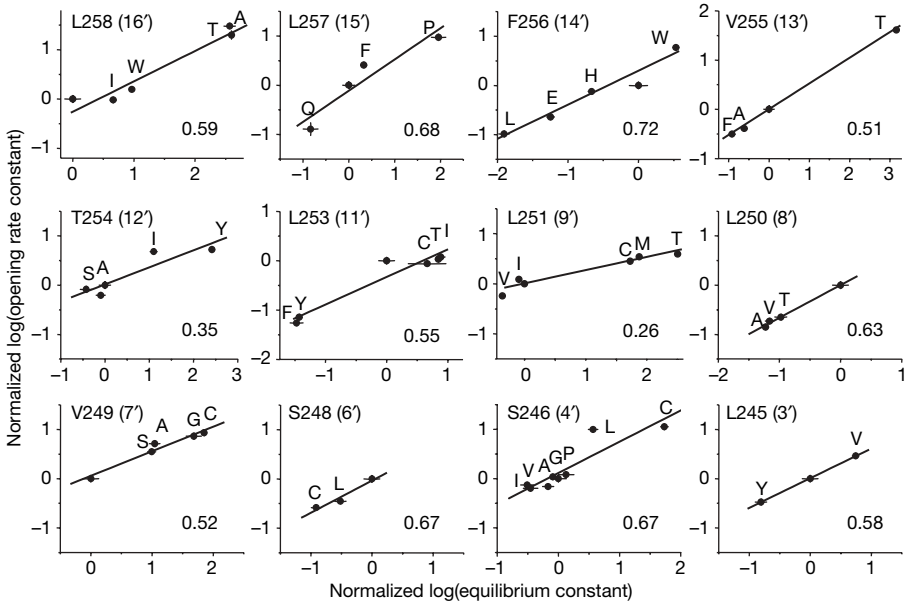


Figure 2 | Φ-value analysis. Plots of normalized log opening rate constant (k_o) against normalized log equilibrium constant ($K_{eq} = k_o/k_c$, where k_c is the closing rate constant) for 12 positions between 16′ and 3′. The wild-type (WT) side chain ($K_{eq} = 0$) is unlabelled; both the rate and equilibrium constants were normalized by the WT values. Three additional αM2 positions are not shown because all constructs exhibited only WT gating kinetics (S10′→A, T, V; I5′→A, G, S, T, W; and T2′→A, C, I). The plots for αM2 positions 27′, 17′ and 9′ were reported previously[17]. The error bars (for both k_o and K_{eq}) are ± s.e.m. (Supplementary Table 1). Φ, the slope of the plot (estimated using the inverse errors on both axes as weights), is shown at lower right of each plot (Supplementary Table 2).

Of the 13 positions between 17′ and 2′ that were sensitive to mutation, six (17′, 16′, 15′, 12′, 9′ and 7′) mainly increased and three (14′, 8′ and 6′) mainly decreased K_{eq}. At four positions (13′, 11′, 4′ and 3′) side-chain substitutions either increased or decreased K_{eq}, by up to ~100-fold. The mutated residues are far from the transmitter binding sites (Fig. 1), and we hypothesize that these changes in K_{eq} arise mainly from parallel changes in the unliganded gating equilibrium constant rather than from changes in the closed/open affinity ratio[21] (Supplementary Fig. 2).

Nine αM2 residues had Φ-values between 0.59 and 0.72 (Fig. 3 and Supplementary Table 2). Included in this group are residues near the cytoplasmic limit (3′), close to the equator (8′), in the upper half (17′) and at the extracellular limit of the helix (27′). We conclude that during di-liganded gating, the movements of most of αM2, from amino to carboxy terminus, are correlated temporally, with $\Phi \approx 0.64$.

Two clearly exceptional residues were 9′ ($\Phi = 0.26$) and 12′ ($\Phi = 0.35$), both of which are near the αM2 equator. The average Φ for these two residues (0.31) is similar to those in the lipid-facing surface of αM3 (0.30)[22], the upper half of δM2 (0.31)[18], the M2 9′ residues of the β- and ε-subunits (0.32)[17] and εM4 (0.33)[23]. Three other αM2 mid-section residues (13′, 11′ and 7′) had Φ-values of 0.51, 0.48 and 0.52, respectively. A weighted k-means cluster analysis of Φ-values for all αM2, αM3 and αM4 residues (Fig. 3a) shows that there are only three populations of Φ value in these domains; it also shows that 13′+11′+7′ and 9′+12′ are distinct from the bulk of αM2 and instead belong to the αM4 and αM3 populations, respectively. We conclude that the gating motions of αM2 probably occur in three steps, with most of the residues throughout the length of the helix moving first ($\Phi = 0.64$), followed by 13′+11′+7′ ($\Phi = 0.52$), followed by 12′+9′ ($\Phi = 0.31$).

The motions of M2 are part of the overall AChR isomerization, in which domains having temporally correlated gating motions ('Φ-blocks') move in sequence to generate a propagated, brownian conformational 'wave' (Fig. 4)[24,25]. In the α-subunit, agonist binding triggers the movement of loops A, B and C in the extracellular domain, which causes a ~10^4-fold increase in the affinity for ACh at each of the two transmitter binding sites ($\Phi \approx 0.93$; purple in

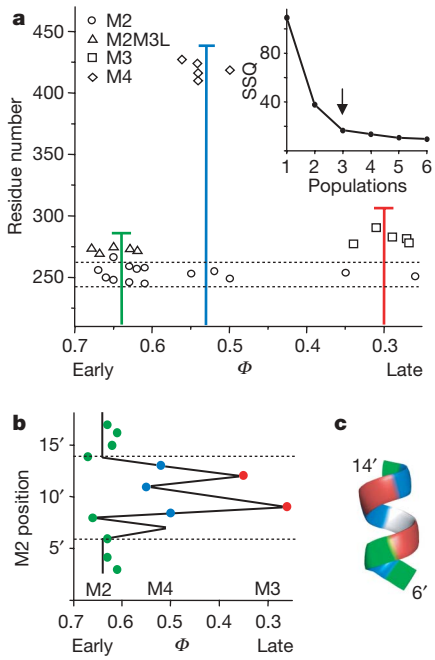


Figure 3 | Distribution of Φ-values. **a**, Statistical analysis of Φ-values for α-subunit residues in M2, the M2-M3 linker (M2M3L) (A. Jha, D. Cadugan, P.P. and A.A., manuscript in preparation), M3[22] and M4[17] (Supplementary Methods). The Φ-values are clustered into three populations (mean ± s.d.): 0.64 ± 0.02 (green), 0.52 ± 0.02 (blue) and 0.30 ± 0.03 (red). Inset, the total sum squared deviation (SSQ) shows that there are three populations of Φ. The αM2 equator has residues in all three populations. **b**, Expansion of the dashed region in **a**. Most of the residues throughout the length of αM2 move relatively early in the gating reaction, but some at the mid-section (7′–13′) move later. No Φ-values are shown for 2′, 5′ and 10′ because mutations here did not significantly change K_{eq}. The order of motion is M2 > M4 > M3. **c**, Cartoon structure (*Torpedo* AChR, PDB accession number 2bg9.pdb) of the dashed region in **b**, coloured by Φ-value (see above; white, <3-fold change in K_{eq}).

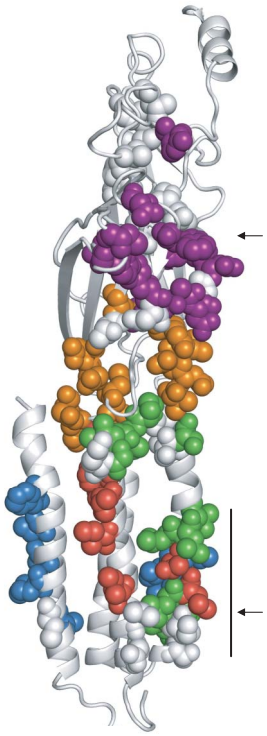


Figure 4 | Map of Φ-values. In the α-subunit shown, Φ-values are clustered spatially and decrease discontinuously between the transmitter binding site and the gate. The vertical line (~20 Å) marks the lumen of the channel, and the upper and lower arrows indicate W149 (at the transmitter binding site) and L251 (at the gate). The cytoplasmic domain (M3–M4 linker) is not shown. The residues that have been examined for Φ-value are shown as spheres[17,21,23,25,26] (A. Jha, D. Cadugan, P.P. and A.A., manuscript in preparation). The colours are according to Φ-value: purple, >0.90; orange, 0.75–0.85; green, 0.59–0.72; blue, 0.48–0.57; red, 0.26–0.35; white, no Φ value (change in K_{eq} < 3-fold). The map suggests that AChR gating is a brownian conformational cascade of domain motions that link an affinity change for agonists at the binding site with a permeability change for ions in the pore (α_ε-subunit, *Torpedo* AChR, PDB accession number 2bg9.pdb).

Fig. 4)[21,25]. This local conformational change then triggers the movement of residues at the base of the extracellular domain, in loop 7 (the Cys-loop) and loop 2 ($\Phi \approx 0.78$; orange)[26]. Next, residues in the M2-M3 linker (A. Jha, D. Cadugan, P.P. and A.A., manuscript in preparation) and throughout most of αM2 move approximately synchronously ($\Phi \approx 0.64$; green), followed in sequence by αM4 and the 13′, 11′ and 7′ residues at the αM2 mid-section ($\Phi \approx 0.52$; blue). Finally, 12′ and 9′ in αM2 move, along with residues in the above-mentioned domains of the α-, β-, δ- and ε-subunits ($\Phi \approx$ 0.30; red), at which point we speculate that ions begin to flow rapidly through the pore. Together, these reversible domain motions provide the framework, if not the structural details, of the AChR channel-opening conformational change.

Two residues have been proposed as key points of energy transfer between the extracellular domain and αM2. Unwin[7] proposed that V46 in loop 2 ($\Phi = 0.78$) interacts with S269 in M2 ($\Phi = 0.64$) to drive opening by a 'pin-into-socket' mechanism. Lummis *et al.*[27] and Lee and Sine[28] suggested that the motion of a proline in the M2-M3 linker (P272 in AChRs) is coupled to the gating motion of M2 by a *cis–trans* isomerization or through interactions with a nearby salt bridge (R209–E45) plus V46. Residues in the M2-M3 linker (including P272) have a Φ-value of ~0.64 (A. Jha, D. Cadugan, P.P. and A.A., manuscript in preparation), which indicates that this region moves synchronously with αM2, after the motion of F135 in the Cys-loop ($\Phi = 0.78$). Also, the pairs F135–P272 and V46–S269 are both at the boundary between the 0.78 and 0.64 Φ-blocks, and mutation of all four of these residues significantly changes K_{eq}.

The propagation of gating energy appears to not be concentrated at a single 'on-off' switch but rather to be spread over at least four Φ-block boundaries ($0.93 \leftrightarrow 0.78 \leftrightarrow 0.64 \leftrightarrow 0.52 \leftrightarrow 0.31$), with the exchange of energy at each boundary occurring at multiple sites. AChR gating (the conformational pathway through the TSE) is neither instantaneous nor smooth but instead is characterized by back-and-forth, brownian motion of nanometre-sized domains, on the 10–100 ns timescale[24].

The Φ-values suggest that the channel opening motions in the membrane are asynchronous. The ends of αM2 move first, then the αM2 centre-flanking regions along with αM4, then the 12′ and 9′ residues of αM2 along with the 9′ residues of βM2, εM2 and αM3. In the δM2 segment, the extracellular end moves about this time, followed by the 9′ residue and the intracellular end[18]. Although our experiments do not address directly whether or not the conformational changes at the middle of M2 regulate ionic conductance, the pattern of Φ-values support the following notion. A 'gate' is located at the middle of M2 (between 9′ and 12′); this gate's structure regulates ionic conductance by movements in the α-subunits that subsequently propagate to the other subunits, rather than by a synchronous, symmetrical motion in all subunits. Previously it was shown that in hybrid AChRs (having only one mutated α-subunit) the two 17′ positions have the same Φ-value but the 9′ positions have different Φ-values (0.0 and 0.32) and, hence, move asynchronously[17]. Because $\Phi = 0.26$ for the double 9′ mutant, we speculate that the difference in Φ for 9′ occurs only in the hybrid constructs and that the gating motions in the M2 double mutants are synchronous.

The first three Φ-blocks (0.93, 0.78 and 0.64) are structurally contiguous, and their movements link the affinity changes at the two transmitter binding sites with structural changes in the region of the gate. However, it is likely that the conductance of the pore does not increase dramatically until the 'late' movements of the equatorial 13′+11′+7′ and 12′+9′ residues of αM2. Higher-resolution C and O structures and a more extensive map of Φ-values may illuminate the chemical details and evolutionary driving forces behind this three-step, doubly locked gating mechanism.

METHODS

Detailed descriptions are in Supplementary Methods. Briefly, mouse α-, β-, δ-and ε-subunits having point mutations in both α-subunits were expressed in HEK cells, and cell-attached recordings were performed using a high agonist concentration ($V_m \approx -100$ mV; 22 °C). Interval durations from selected single-channel clusters were fitted directly to estimate opening and closing rate constants. Φ is the slope of a log–log plot of opening rate-constant versus equilibrium constant for a series of mutations of each residue.

There are three plausible, non-exclusive interpretations for the physical meaning of Φ. (1) If there are many conformational pathways connecting C and O, then Φ is a function of the probabilities of traversing each of these, and reflects structural heterogeneity of the TSE. (2) If the conformation of an individual side chain at the TSE is not just either 'closed' or 'open' but can also be somewhere in between, then Φ reflects the fractional extent of reaction at the residue level (the closed- versus open-like structure of the perturbed side chain). (3) If domains sequentially and instantaneously flip conformation and in the TSE there is a mixture (with some domains 'closed' and others 'open'), then Φ reflects the fractional extent of reaction at the domain level (the closed- versus open-like structure of the protein when the perturbed side chain flips). All of these interpretations are consistent with an apparent two-state reaction.

The AChR is large, and undoubtedly there are metastable intermediates between stable C and stable O. Experiments show that the net dynamics across the TSE are slow (~μs^{-1})[29] and that the map of Φ is organized. There are spatially contiguous groups of residues that have approximately the same Φ-value, and there is an approximately longitudinal gradient in Φ that decreases discontinuously between the transmitter binding site and the gate (Fig. 4)[23,25]. A two-pathway mechanism ($\Phi = 0$ and 1) cannot explain some fractional Φ-values[18], and only the domain-level interpretation predicts such an organized Φ map. We hypothesize that Φ reflects the fractional extent of reaction at the domain level and, hence, gives the relative timing of the perturbed residue's motion within the overall reaction cascade[24].

Received 20 December 2006; accepted 28 February 2007.

1. Edelstein, S. & Changeux, J.-P. Allosteric transitions of the acetylcholine receptor. *Adv. Protein Chem.* **51,** 121–184 (1998).
2. Karlin, A. Emerging structure of the nicotinic acetylcholine receptors. *Nature Rev. Neurosci.* **3,** 102–114 (2002).
3. Lester, H. A., Dibas, M. I., Dahan, D. S., Leite, J. F. & Dougherty, D. A. Cys-loop receptors: New twists and turns. *Trends Neurosci.* **27,** 329–336 (2004).
4. Sine, S. M. & Engel, A. G. Recent advances in Cys-loop receptor structure and function. *Nature* **440,** 448–455 (2006).
5. Zhou, Y., Pearson, J. E. & Auerbach, A. Φ-Value analysis of a linear, sequential reaction mechanism: Theory and application to ion channel gating. *Biophys. J.* **89,** 3680–3685 (2005).
6. Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423,** 949–955 (2003).
7. Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346,** 967–989 (2005).
8. Akabas, M. H., Kaufmann, C., Archdeacon, P. & Karlin, A. Identification of acetylcholine receptor channel-lining residues in the entire M2 segment of the α-subunit. *Neuron* **13,** 919–927 (1994).
9. Wilson, G. G. & Karlin, A. The location of the gate in the acetylcholine receptor channel. *Neuron* **20,** 1269–1281 (1998).
10. Paas, Y. *et al.* Pore conformations and gating mechanism of a Cys-loop receptor. *Proc. Natl Acad. Sci. USA* **102,** 15877–15882 (2005).
11. Bali, M. & Akabas, H. The location of closed channel gate in Cys-loop receptor channels. *J. Gen. Physiol.* (in the press).
12. Panicker, S., Cruz, H., Arrabit, C. & Slesinger, P. A. Evidence for a centrally located gate in the pore of a serotonin-gated ion channel. *J. Neurosci.* **22,** 1629–1639 (2002).
13. Beckstein, O. & Sansom, M. S. P. A hydrophobic gate in an ion channel: The closed state of the nicotinic acetylcholine receptor. *Phys. Biol.* **3,** 147–159 (2006).
14. Labarca, C. *et al.* Channel gating governed symmetrically by conserved leucine residues in the M2 domain of nicotinic receptors. *Nature* **376,** 514–516 (1995).
15. Pascual, J. M. & Karlin, A. State-dependent accessibility and electrostatic potential in the channel of the acetylcholine receptor. inferences from rates of reaction of thiosulfonates with substituted cysteines in the M2 segment of the α subunit. *J. Gen. Physiol.* **111,** 717–739 (1998).
16. Leffler, J. E. & Grunwald, E. *Rates and Equilibria of Organic Reactions as Treated by Statistical, Thermodynamic, and Extrathermodynamic Methods* (Wiley, New York, 1963).
17. Mitra, A., Cymes, G. D. & Auerbach, A. Dynamics of the acetylcholine receptor pore at the gating transition state. *Proc. Natl Acad. Sci. USA* **102,** 15069–15074 (2005).
18. Cymes, G. D., Grosman, C. & Auerbach, A. Structure of the transition state of gating in the acetylcholinereceptor channel pore: A Φ-value analysis. *Biochemistry* **41,** 5548–5555 (2002).
19. Villarroel, A., Herlitze, S., Koenen, M. & Sakmann, B. Location of a threonine residue in the α-subunit M2 transmembrane segment that determines the ion flow through the acetylcholine receptor channel. *Proc. R. Soc. Lond. B* **243,** 69–74 (1991).
20. Zhang, H. & Karlin, A. Contribution of the beta subunit M2 segment to the ion-conducting pathway of the acetylcholine receptor. *Biochemistry* **37,** 7952–7964 (1998).
21. Chakrapani, S., Bailey, T. D. & Auerbach, A. The role of loop 5 in acetylcholine receptor channel gating. *J. Gen. Physiol.* **122,** 521–539 (2003).
22. Cadugan, D. J. & Auerbach, A. Conformational dynamics of the αM3 transmembrane helix during acetylcholine receptor channel gating. *Biophys. J.* (in the press).
23. Mitra, A., Bailey, T. D. & Auerbach, A. L. Structural dynamics of the M4 transmembrane segment during acetylcholine receptor gating. *Structure* **12,** 1909–1918 (2004).
24. Auerbach, A. Gating of acetylcholine receptor channels: Brownian motion across a broad transition state. *Proc. Natl Acad. Sci. USA* **102,** 1408–1412 (2005).
25. Grosman, C., Zhou, M. & Auerbach, A. Mapping the conformational wave of acetylcholine receptor channel gating. *Nature* **403,** 773–776 (2000).
26. Chakrapani, S., Bailey, T. D. & Auerbach, A. Gating dynamics of the acetylcholine receptor extracellular domain. *J. Gen. Physiol.* **123,** 341–356 (2004).
27. Lummis, S. C. R. *et al.* Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* **438,** 248–252 (2005).
28. Lee, W. Y. & Sine, S. M. Principal pathway coupling agonist binding to channel gating in nicotinic receptors. *Nature* **438,** 243–247 (2005).
29. Chakrapani, S. & Auerbach, A. A speed limit for conformational change of an allosteric membrane protein. *Proc. Natl Acad. Sci. USA* **102,** 87–92 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Grosman, P. Gottlieb and F. Sachs for discussions, and C. Nicolai, B. Steidl, M. Merritt and M. Teeling for technical assistance. This work was supported by the NIH.

Author Contributions All of the experiments and data analyses shown in Fig. 2 were performed by P.P. except for those for positions 4′ and 9′, which were performed by A.M.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.A. (auerbach@buffalo.edu).

LETTERS

Regioselective one-pot protection of carbohydrates

Cheng-Chung Wang[1,4], Jinq-Chyi Lee[1], Shun-Yuan Luo[1], Suvarn S. Kulkarni[2,3], Yu-Wen Huang[1], Chia-Chen Lee[1], Ken-Lien Chang[1] & Shang-Cheng Hung[1,2,3,4]

Carbohydrates are involved in a wide range of biological processes[1–4]. These structurally diverse compounds are more complex than other biological polymers, and are often present as heterogeneous mixtures in nature. The chemical synthesis of carbohydrates is one way to obtain pure oligosaccharides, but it is hampered by difficulties associated with the regioselective protection of polyhydroxyls and challenges related to the stereoselective assembly of glycosidic linkages[5–14]. Here we describe a combinatorial, and highly-regioselective, method that can be used to protect individual hydroxy groups of a monosaccharide. This approach can be used to install an orthogonal protecting group pattern in a single reaction vessel (a 'one-pot' reaction), which removes the need to carry out the time-consuming isolation and purification of intermediates. Hundreds of building blocks have been efficiently prepared starting from D-glucose, and the iterative coupling of these building blocks enabled us to assemble β-1,6-glucans and a library of oligosaccharides based on the influenza-virus-binding trisaccharide.

Oligosaccharides and glycoconjugates have significant roles in a diverse set of biological processes, including viral and bacterial infections, cell growth and proliferation, cell–cell communication, as well as immunoresponse[1]. Their structural diversity, which allows them

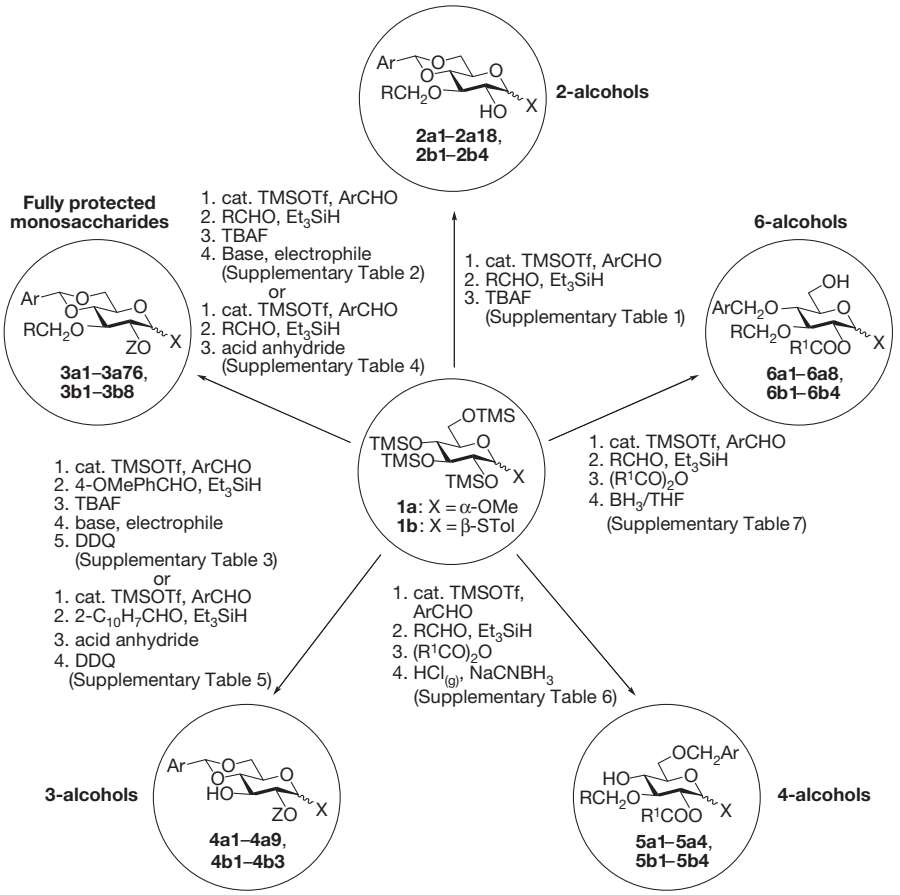


Figure 1 | One-pot protection of carbohydrates. A combinatorial, regioselective, orthogonal, and one-pot protection of the per-*O*-silylated α-glucosides **1a** and β-thioglucoside **1b** to produce fully protected monosaccharides, 2-alcohols, 3-alcohols, 4-alcohols, and 6-alcohols. Starting from **1a** and **1b**, a panoply of diversely protected sugar blocks is rapidly prepared. These synthons can readily be used as glycosyl acceptors and donors in the assembly of oligosaccharides. The procedure uses a single catalyst, TMSOTf, giving selectively one isomer at each step, thus allowing multi-step transformation in a single flask. Reagents: ArCHO, aryl aldehyde; DDQ, 2,3-dichloro-5,6-dicyano-1,4-benzoquinone; BH_3, borane; THF, tetrahydrofuran; cat., catalytic amount; Me, methyl; Ph, phenyl; Et, ethyl; $HCl_{(g)}$, gaseous HCl; STol, thiotoluenyl; Ar, aryl.

[1]Department of Chemistry, National Tsing Hua University, Hsinchu 300, Taiwan. [2]Genomics Research Center, [3]Institute of Chemistry, [4]Chemical Biology and Molecular Biophysics Program, Taiwan International Graduate Program, Academia Sinica, Taipei 115, Taiwan.

to encode information required for specific molecular recognition, and determine the post-translational modification of proteins, is much more complex than that of proteins and nucleic acids. The preparation[2–4] of oligosaccharides is more difficult than that of proteins and nucleic acids because no regio- and stereochemical issues are involved in the sequential coupling steps for the construction of amide or phosphate bonds, respectively. The biggest challenge in carbohydrate synthesis is not only the stereoselective glycosylation[5–14], but also the preparation of selectively protected monosaccharide units, one with a strategically positioned free hydroxy group (a nucleophilic acceptor) and one bearing a labile leaving group at the anomeric carbon that acts as a glycosyl donor in the ensuing glycosylation reaction. It is also necessary to install suitable protecting groups on the remaining hydroxyls, for tuning of the overall electronic properties of the donors and acceptors so as to 'match' the donor–acceptor pair and also for further deprotection and glycosylation or functional-group modifications.

A traditional chemical approach for oligosaccharide preparation involving multi-step protection and glycosylation is schematically illustrated in Supplementary Fig. 1. In a straightforward synthesis of carbohydrates (Supplementary Fig. 2), we envisioned that the 2-, 3-, 4-, and 6-hydroxy groups on a monosaccharide bearing an anomeric group (X = α- or β-OR, -SR and -SeR) could be separately protected via a combinatorial, regioselective, orthogonal, and sequential one-pot procedure to obtain either the fully protected monosaccharides or the individual alcohols. These building blocks can be rapidly coupled through the one-pot glycosylation process to yield various oligosaccharide skeletons.

For the successful development of this one-pot protection scheme, the choice of the protecting groups was crucial. We chose substituted and unsubstituted benzyl ethers (Supplementary Fig. 3), which have a larger family than esters or silyl ethers, act as orthogonal protecting groups and can be cleaved advantageously by means of reagent combination[15–20]. Strategically, the O2-protecting group of a monosaccharide unit that is used as a glycosyl donor plays a pivotal role in the stereocontrol of glycosylation. Benzyl-type ethers at the C2 carbon-atom position of the donor in general favour the formation of axially oriented glycosidic bonds by virtue of the anomeric effect, whereas installation of 1,2-*trans*-linkage is usually achieved by neighbouring-group participation of an ester-type protecting group at C2. Thus we needed both benzyl-type groups and the incorporation of acyl groups at oxygen atom O2 in the one-pot protection endeavour.

To test this strategy, D-glucose was selected as a substrate for our studies. The D-glucopyranosyl 2,3,4,6-tetraols cannot dissolve in most organic solvents; their per-*O*-silylation not only provided the requisite *O*-trimethylsilylated functionalities for regioselective protection, but also dramatically improved the solubility of unprotected monosaccharides. The corresponding 2,3,4,6-tetra-*O*-TMS ethers **1a** and **1b** (Supplementary Fig. 4) were prepared from methyl α-D-glucopyranoside and *p*-toluenyl β-D-thioglucopyranoside in near-quantitative yields, respectively. The reaction conditions and the results of their differential one-pot protection, to generate a vast array of building blocks, are summarized in Fig. 1. Our general protocols involved four steps: (1) selective protection of oxygen atoms O4 and O6 as an arylidene acetal followed by highly regioselective reductive arylmethylation at oxygen O3 to produce the 2-alcohols, (2) O4,O6-arylidenation, O3-arylmethylation, and subsequent etherification or acylation at O2 to get the fully protected monosaccharides, (3) O4,O6-arylidenation, *p*-methoxybenzyl or 2-naphthylmethyl protection at O3, O2-etherification or O2-acylation, and removal of *p*-methoxybenzyl or 2-naphthylmethyl group to yield the 3-alcohols, and (4) O4,O6-arylidenation, O3-arylmethylation, O2-acylation, and regioselective ring opening of arylidene acetals at O4 and O6 to provide the 4-alcohols and 6-alcohols, respectively. The novelty of this approach lies in tuning of the reaction conditions so as to generate a single regioisomer at each stage that allows sequential addition of reagents in the same pot.

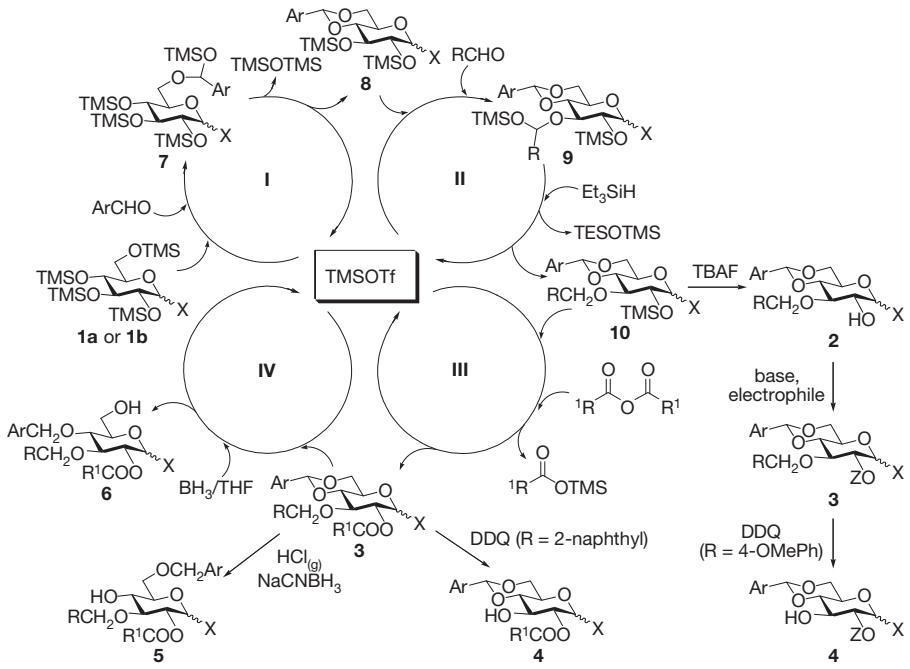


Figure 2 | Proposed mechanism for the TMSOTf-catalysed one-pot protection of carbohydrates. Catalytic cycle I shows TMSOTf-catalysed O4,O6-arylidenation of **1a** or **1b** with an aryl aldehyde. Catalytic cycle II shows TMSOTf-catalysed regioselective Et_3SiH-reductive O3-etherification of the acetal **8** with the second aryl aldehyde. O2-desilylation of compound **10** provides the 2-alcohols **2** and the freed hydroxy group can be further protected to give the fully protected monosaccharide **3**. When R is a 4-methoxyphenyl group, addition of DDQ to compound **3** in the same pot yields the corresponding 3-alcohol **4**. Catalytic cycle III shows TMSOTf-catalysed O2-acylation of **10** with an acid anhydride. O3-cleavage (when R = 2-naphthyl) and O4-ring opening of compound **3** in the same pot produces the 3-alcohol **4** and 4-alcohol **5**, respectively. Catalytic cycle IV shows TMSOTf-catalysed regioselective borane-reductive O6-ring opening of the arylidene acetal **3**. Reagents: TMS, trimethylsilyl; BH_3, borane; $NaCNBH_3$, sodium cyanoborohydride; TESOTMS, triethylsilyloxytrimethylsilane; R^1 is a methyl or phenyl group.

Accordingly, compound **1a** or **1b** was first treated with one equivalent of aryl aldehyde in the presence of trimethylsilyl trifluoromethanesulphonate (TMSOTf) as a catalyst at ice-bath temperature. After ring formation of arylidene acetal at oxygens O4 and O6 was complete, a different aryl aldehyde or another equivalent of the same aldehyde was added to the mixture at subzero temperature (–86 °C) followed by treatment with triethylsilane (Et_3SiH)[21,22]. The reaction was exceptionally clean, giving a single O3-ether; the excellent regioselectivity may conceivably be attributed to the higher nucleophilicity of the 3-oxygen atom[23]. The mixture was then treated with tetra-*n*-butylammonium fluoride (TBAF) to remove the TMS group at O2, and the corresponding 2-alcohols **2a1**–**2a18** and **2b1**–**2b4**, as listed in Supplementary Table 1, were obtained in high overall yields, respectively.

In another set of experiments, the first three steps were repeated, and the appropriate electrophile was added under basic conditions to get the fully protected glycosides **3a1**–**3a63** and **3b1**–**3b8** in good yields and in a one-pot manner (Supplementary Table 2), including the introduction of various benzyl, allyl and acyl type groups at the O2 position. For the preparation of 3-alcohols, *p*-methoxybenzyl was used as a temporary protecting group at O3. The same operation as described for the fully protected glycosides was repeated, after which DDQ was added to the reaction mixture, and the 3-OH compounds **4a1**–**4a5** and **4b1**–**4b2** were obtained (Supplementary Table 3). We realized that O2-acylation of the 2-*O*-trimethylsilyl (2-OTMS) intermediate, generated after the first two steps, could be achieved under the prevailing acidic conditions of TMSOTf as well, using acid anhydrides as reagents to produce the fully protected monosaccharides (Supplementary Table 4). Likewise, the 3-OH derivatives **4a1**, **4a6**–**4a9** and **4b3** were isolated in good yields via subsequent oxidative cleavage of the 2-naphthylmethyl group at oxygen O3 with DDQ in one flask (Supplementary Table 5). The 4-alcohols (Supplementary Table 6, **5a1**–**5a4** and **5b1**–**5b4**) and 6-alcohols (Supplementary Table 7, **6a1**–**6a8** and **6b1**–**6b4**) were in turn accessed via regioselective O4- and O6-ring opening of arylidene acetals, formed by tandem transformations, O4,O6-arylidenation, O3-arylmethylation, and O2-acylation, on **1a** or **1b**, in the same pot using $HCl_{(g)}$/$NaBH_3CN$ and BH_3/THF as reductants, respectively. Thus, D-glucopyranosides can be efficiently converted into various glycosyl donors and acceptors bearing chemically differentiable protective groups. We note that this method worked equally well at a larger scale (see Methods). Moreover, our preliminary results indicate that the protocol can be adopted for other sugars as well (Supplementary Fig. 5).

The catalytic role played by TMSOTf in these reactions is intriguing. The results can be rationalized on the basis of catalytic cycles, as outlined in Fig. 2. In the first catalytic cycle, reaction of **1a** or **1b** with an aryl aldehyde initially forms a TMS-acetal **7**, at the most accessible O6. Owing to the proximity of the 4-OTMS group, the intermediate **7** is transformed into an arylidene acetal **8** with the liberation of TMSOTMS. TMSOTf then catalyses regioselective O3-etherification of **8** with the second aryl aldehyde and Et_3SiH to produce the 2-OTMS ether **10** via the intermediate TMS-acetal **9** in the second catalytic cycle. The TMS group at the O2 position of compound **10** could be cleaved by TBAF to provide the 2-alcohol **2**, and the freed hydroxy group can be further manipulated under basic conditions to yield the fully protected compound **3**. When R is a 4-methoxyphenyl group, extra-addition of DDQ to **3** in the same pot leads to the 3-alcohol **4**. Without quenching the 2-OTMS ether **10**, TMSOTf can concomitantly catalyse acylation of **10** with an acid anhydride under acidic conditions to give the fully protected ester **3** (catalytic cycle III). The 4,6-*O*-arylidene acetal can be selectively opened at O6 by further treatment of **3** with BH_3/THF in the same pot under the prevalent conditions in the catalytic cycle IV to afford the 6-alcohol **6**. In addition, regioselective O3-cleavage (when R = 2-naphthyl group) and O4-cleavage of **3** using DDQ and $HCl_{(g)}$/$NaCNBH_3$ in the same pot give the corresponding 3-OH compound **4** and 4-alcohol **5**, respectively.

With these diverse building blocks in hand, we further investigated iterative one-pot glycosylation[24] for the assembly of β-1,6-glucans with various chain lengths. A fully protected thioglucoside **3b2**, a thioglucosyl 6-alcohol **6b2**, and a methyl glucosyl 6-alcohol **6a2** were used as the starting, elongation and termination units, respectively, to rapidly construct the di-, tri-, tetra- and pentasaccharides in a one-pot manner and in good overall yields (Supplementary Fig. 6).

The H5N1 avian influenza virus continues to spread worldwide, with implications for public health: see www.who.int/en. Direct bird-to-human transmission is considered a potential threat because the introduction of viral mutations can cause a switch in receptor specificity from D-sialic acids attached to D-galactose in α2,3-linkages (avian) to α2,6-linkages (human)[25]. A trisaccharide unit—SAα2,6Galβ1,4GlcNHAc (Fig. 3a)—was identified as a host-cell receptor for haemagglutinin binding during the initial step of viral infection[26,27]. We could find no reports of one-pot glycosylation using the sialyl donor as the starting sugar unit via chemical approach.

To address this problem, we present here, for the first time, the one-pot synthesis of this trisaccharide and its derivatives (Fig. 3b, c). The D-galactopyranosyl 6-alcohol **12** was prepared from the tetra-*O*-TMS ether **11** in 56% yield via the one-pot protection strategy, involving 4,6-*O*-benzylidenation, 2,3-di-*O*-benzoylation and O6-ring

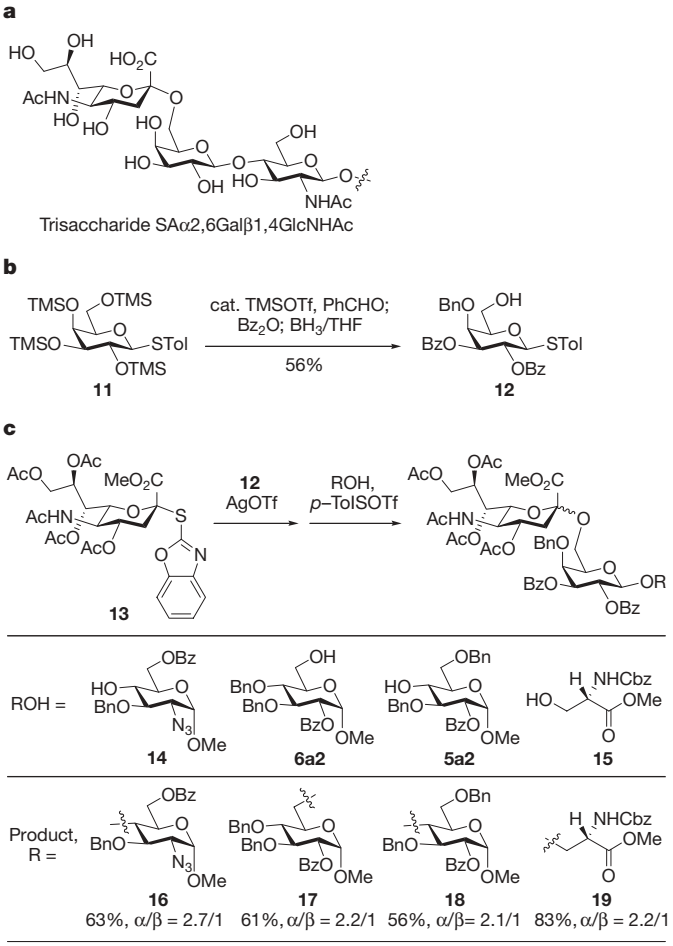


Figure 3 | One-pot synthesis of influenza virus-binding trisaccharide library using the sialyl donor 13 as the starting sugar unit. a, Influenza virus-binding trisaccharide of human host cell surface. **b**, One-pot synthesis of the galactosyl 6-alcohol **12** from the tetra-*O*-TMS ether **11**. **c**, One-pot glycosylation of the sialyl donor **13** and **12** with various alcohols. Yield is shown as a percentage. Reagents: AgOTf, silver trifluoromethanesulphonate; Bn, benzyl; Ac, acetyl; Bz, benzoyl (PhCO); Cbz, benzyloxycarbonyl; Me, methyl. Diastereomers α and β, the glycosidic bonds at the anomeric carbon atom of sialic acid orient toward the equatorial and axial positions, respectively.

opening. All the initial attempts for the coupling of **12** with various sialyl donors bearing different anomeric leaving groups and protecting groups failed to yield the desired disaccharides, except when we used compound **13** (ref. 28) in a mixed solvent of CH_2Cl_2 and Et_2O that led to the α-isomer as the major product. AgOTf-activated assembly of **13** and **12**, followed by *p*-toluenylsulphenyl trifluoromethanesulphonate (*p*-TolSOTf)-promoted coupling with the alcohols **14** (ref. 29), **6a2**, **5a2**, and **15** in a one-pot manner, yielded the expected trisaccharides **16**, **17**, and **18** as well as the glycoconjugate **19**, respectively. These products and their β-sialyl isomers can be individually purified by column chromatography and the stereochemistry of each diastereoisomer was well-characterized through analysis of its ^{1}H, ^{13}C, distortionless enhancement by polarization transfer (DEPT), ^{1}H–^{1}H correlation spectroscopy (COSY), ^{1}H–^{13}C COSY, and Nuclear Overhauser effect spectroscopy (NOESY) nuclear magnetic resonance spectra.

In conclusion, we have developed a highly regioselective, combinatorial, orthogonal and one-pot protection of monosaccharides. Conceptually, this is the first approach (to our knowledge) to discriminate sugar polyols in a single flask, using a single catalyst. Via the TMSOTf-catalysed reaction, the desired building block with differential protective group pattern can be accessed easily after a work-up and single purification on large scales. The reaction conditions are optimized for D-glucopyranosides but should be applicable to other sugars as well, as is indicated by our preliminary studies on D-mannoside, D-galactoside and 2-azido-2-deoxy-D-glucoside. This new protocol, in conjunction with one-pot glycosylation, should help to expedite the overall synthetic process and reduce the labour involved in saccharide preparation.

METHODS

Synthesis of 2b1**.** To obtain **2b1** (*p*-methylphenyl 3-*O*-benzyl-4,6-*O*-benzylidene-β-D-thioglucopyranoside), we added TMSOTf (1.9 ml, 10.4 mmol) at 0 °C under nitrogen to a solution of compound **1b** (39.9 g, 69.4 mmol), PhCHO (7.2 ml, 70.8 mmol), and freshly dried 3 Å molecular sieves (20 g) in dichloromethane (300 ml). After stirring for 2 h, the reaction flask was cooled down to –78 °C, and Et_3SiH (11.1 ml, 69.4 mmol), PhCHO (7.2 ml, 70.8 mmol) and TMSOTf (1.2 mL, 6.9 mmol) were sequentially added to the solution; the mixture was continuously stirred at the same temperature overnight. A 1 M solution of TBAF in THF (139 ml, 139 mmol) was added to the mixture, the reaction flask was gradually warmed up to room temperature, and the resulting solution was kept stirring for 3 h. The whole mixture was filtered through a pad of celite, and the filtrate was washed with water (100 ml). The aqueous layer was extracted by EtOAc (3 × 50 ml), and the combined organic layers were washed with brine, dried over anhydrous $MgSO_4$, filtered, and concentrated under vacuum. Recrystallization of this residue in EtOH yielded the desired 2-alcohol **2b1** (26.2 g, 81%). For the full protocols see the Supplementary Information.

Received 12 October 2006; accepted 2 March 2007.

1. Varki, A., *et al.* (eds) *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, New York, 1999).
2. Ernst, B., Hart, G. W. & Sinaÿ, P. (eds) *Carbohydrates in Chemistry and Biology* Vol. 1 (Wiley, Weinheim, 2000).
3. Bertozzi, C. R. & Kiessling, L. L. Chemical glycobiology. *Science* **291,** 2357–2364 (2001).
4. Schofield, L. *et al.* Synthetic GPI as a candidate anti-toxin vaccine in model of malaria. *Nature* **418,** 785–789 (2002).
5. Yamada, H., Harada, T., Miyazaki, H. & Takahashi, T. One-pot sequential glycosylation: a new method for the synthesis of oligosaccharides. *Tetrahedron Lett.* **35,** 3979–3982 (1994).
6. Douglas, N. L., Ley, S. V., Lücking, U. & Warriner, S. L. Tuning glycoside reactivity: New tool for efficient oligosaccharide synthesis. *J. Chem. Soc. Perkin Trans. I* 51–65 (1998).
7. Zhang, Z. *et al.* Programmable one-pot oligosaccharide synthesis. *J. Am. Chem. Soc.* **121,** 734–753 (1999).
8. Sears, P. & Wong, C.-H. Toward automated synthesis of oligosaccharides and glycoproteins. *Science* **291,** 2344–2350 (2001).
9. Plante, O. J., Palmacci, E. R. & Seeberger, P. H. Automated solid-phase synthesis of oligosaccharides. *Science* **291,** 1523–1527 (2001).
10. Danishefsky, S. J., McClure, K. F., Randolph, J. T. & Ruggeri, R. B. A strategy for the solid-phase synthesis of oligosaccharides. *Science* **260,** 1307–1309 (1993).
11. Kim, J.-H., Yang, H., Park, J. & Boons, G.-J. A general strategy for stereoselective glycosylations. *J. Am. Chem. Soc.* **127,** 12090–12097 (2005).
12. Flitsch, S. L. Glycosylation with a twist. *Nature* **437,** 201–202 (2005).
13. Demchenko, A. V. Stereoselective chemical 1,2-*cis*-*O*-glycosylation: from ''sugar ray'' to modern techniques of the 21st century. *Synlett* 1225–1240 (2003).
14. Pellissier, H. Use of *O*-glycosylation in total synthesis. *Tetrahedron* **61,** 2947–2993 (2005).
15. Kocienski, P. J. *Protecting Groups* 3rd edn (Georg Thieme, Stuttgart, 2005).
16. Wuts, P. G. M. *Greene's Protective Groups in Organic Synthesis* 4th edn (John Wiley & Sons, New York, 2007).
17. Wright, J. A., Yu, J. & Spencer, J. B. Sequential removal of the benzyl-type protecting groups PMB and NAP by oxidative cleavage using CAN and DDQ. *Tetrahedron Lett.* **42,** 4033–4036 (2001).
18. Plante, O. J., Buchwald, S. L. & Seeberger, P. H. Halobenzyl ethers as protecting groups for organic synthesis. *J. Am. Chem. Soc.* **122,** 7148–7149 (2000).
19. Jobron, L. & Hindsgaul, O. Novel *para*-substituted benzyl ethers for hydroxyl group protection. *J. Am. Chem. Soc.* **121,** 5835–5836 (1999).
20. Shie, C.-R. *et al.* $Cu(OTf)_2$ as an efficient and dual-purpose catalyst in the regioselective reductive ring opening of benzylidene acetals. *Angew. Chem. Int. Edn* **44,** 1665–1668 (2005).
21. Tsunoda, T., Suzuki, M. & Noyori, R. A facile procedure for acetalization under aprotic conditions. *Tetrahedron Lett.* **21,** 1357–1358 (1980).
22. Hatakeyama, S. *et al.* Efficient reductive etherification of carbonyl compounds with alkoxytrimethylsilanes. *Tetrahedron Lett.* **35,** 4367–4370 (1994).
23. Wang, C.-C. *et al.* Synthesis of biologically potent α1,2-linked disaccharide derivatives via regioselective one-pot protection glycosylation. *Angew. Chem. Int. Edn* **41,** 2360–2362 (2002).
24. Huang, X., Huang, L., Wang, H. & Ye, X.-S. Iterative one-pot synthesis of oligosaccharides. *Angew. Chem. Int. Edn* **43,** 5221–5224 (2004).
25. Parrish, C. R. & Kawaoka, Y. The origins of new pandemic viruses: The acquisition of new host ranges by canine parvovirus and influenza A viruses. *Annu. Rev. Microbiol.* **59,** 553–586 (2005).
26. Gamblin, S. J. *et al.* The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **303,** 1838–1842 (2004).
27. Stevens, J. *et al.* Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science* **303,** 1866–1870 (2004).
28. De Meo, C. & Parker, O. Thiomidoyl approach to the synthesis of α-sialosides. *Tetrahedron Asymm.* **16,** 303–307 (2005).
29. Lee, J.-C. *et al.* Synthesis of Heparin Oligosaccharides. *J. Am. Chem. Soc.* **126,** 476–477 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C.-C. Liao and H.-J. Liu for their discussions. This work was supported by the National Science Council of Taiwan and the Academia Sinica. S.S.K. thanks Academia Sinica for a postdoctoral fellowship.

Author Contributions S.-C.H. conceived the idea of one-pot protection, supervised students to carry out the experiments, and outlined the figures in the manuscript. C.-C.W. initiated the extensive work on the protection of methyl glucopyranoside and carried out the synthesis of β-1,6-glucans and haemagglutinin-binding trisaccharides. S.-Y.L. synthesized additional methyl glucopyranosides. J.-C.L. standardized the initial reactions on thioglycosides. S.S.K. synthesized more thioglycosides and wrote the manuscript. Y.-W.H., C.-C.L. and K.-L.C. contributed a representative example each of D-galactoside, D-mannoside, and 2-azido-2-deoxy-D-glucoside, respectively.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.-C.H. (hung@mx.nthu.edu.tw).

LETTERS

TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity

Michaela U. Gack[1,2], Young C. Shin[1], Chul-Hyun Joo[1,3], Tomohiko Urano[4,5], Chengyu Liang[1], Lijun Sun[6], Osamu Takeuchi[7], Shizuo Akira[7], Zhijian Chen[6], Satoshi Inoue[4,5] & Jae U. Jung[1]

Retinoic-acid-inducible gene-I (RIG-I; also called DDX58) is a cytosolic viral RNA receptor that interacts with MAVS (also called VISA, IPS-1 or Cardif) to induce type I interferon-mediated host protective innate immunity against viral infection[1–6]. Furthermore, members of the tripartite motif (TRIM) protein family, which contain a cluster of a RING-finger domain, a B box/coiled-coil domain and a SPRY domain, are involved in various cellular processes, including cell proliferation and antiviral activity[7]. Here we report that the amino-terminal caspase recruitment domains (CARDs) of RIG-I undergo robust ubiquitination induced by TRIM25 in mammalian cells. The carboxy-terminal SPRY domain of TRIM25 interacts with the N-terminal CARDs of RIG-I; this interaction effectively delivers the Lys 63-linked ubiquitin moiety to the N-terminal CARDs of RIG-I, resulting in a marked increase in RIG-I downstream signalling activity. The Lys 172 residue of RIG-I is critical for efficient TRIM25-mediated ubiquitination and for MAVS binding, as well as the ability of RIG-I to induce antiviral signal transduction. Furthermore, gene targeting demonstrates that TRIM25 is essential not only for RIG-I ubiquitination but also for RIG-I-mediated interferon-β production and antiviral activity in response to RNA virus infection. Thus, we demonstrate that TRIM25 E3 ubiquitin ligase induces the Lys 63-linked ubiquitination of RIG-I, which is crucial for the cytosolic RIG-I signalling pathway to elicit host antiviral innate immunity.

A recent series of studies has identified RIG-I and melanoma differentiation-associated gene 5 (MDA5; also called IFIH1) as cytosolic receptors for viral double-stranded RNA and 5′ triphosphate RNA[2,4,6]. RIG-I and MDA5 belong to the DExD/H box RNA helicase family, the members of which contain two caspase recruitment domains (2CARD) in the N-terminal region and a potential ATP-dependent RNA helicase activity in the C-terminal region[8,9]. To decipher the cytosolic RIG-I-mediated antiviral signalling pathway, we attempted to identify cellular proteins associated with the N-terminal 2CARD of RIG-I and MDA5 using mammalian glutathione *S*-transferase (GST) fusion constructs. Polypeptides with apparent molecular masses of 52, 60 and 68 kDa were present specifically in the GST–RIG-I(2CARD) complex but not in the GST–MDA5(2CARD) complex or with GST alone (Fig. 1a). Notably, mass spectrometry and immunoblotting showed that these polypeptides were exclusively identified as ubiquitinated forms of GST–RIG-I(2CARD) (Supplementary Fig. 1a). To confirm RIG-I ubiquitination, HEK293T cells were co-transfected with Flag-tagged full-length RIG-I or a RIG-I mutant in which the 2CARD had been deleted (RIG-I(Δ2CARD)) together with haemagglutinin (HA)-tagged ubiquitin. RIG-I, but not RIG-I(Δ2CARD), was extensively ubiquitinated (Fig. 1b and Supplementary Fig. 1b). In addition, anti-HA immunoblotting detected ubiquitinated Flag-tagged RIG-I as multiple species with apparent molecular masses of 120–150 kDa, significantly larger than unmodified Flag–RIG-I (Fig. 1c, top left panel). Furthermore, Sendai virus infection and/or interferon (IFN)-β treatment resulted in the markedly increased ubiquitination of endogenously or exogenously expressed RIG-I (Fig. 1c and Supplementary Fig. 1c). These results indicate that RIG-I undergoes robust ubiquitination at its N-terminal 2CARD and that this ubiquitination apparently increases on viral infection.

To dissect further the ubiquitination of the 2CARD of RIG-I, which contains 18 lysine residues, the *in vivo* ubiquitinated forms of N-terminal GST-fused and C-terminal Flag-tagged RIG-I(2CARD) were purified and analysed by multi-dimensional liquid chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS; Fig. 1a, d, bands 1–3). Both GST–RIG-I(2CARD) and Flag–RIG-I(2CARD) carried the ubiquitin peptides at Lys 99, 169, 172, 181, 190 or 193 (Fig. 1d). Additional mass spectrometry analysis showed that band 2 and 3 fragments carried the unique, branched Gly-Gly signature peptides primarily with the ubiquitin Lys 63 linkage (Fig. 1d). Furthermore, GST–RIG-I(2CARD) and full-length RIG-I were strongly ubiquitinated when HA-tagged wild-type ubiquitin or a K48R ubiquitin mutant was expressed, whereas their ubiquitination was significantly reduced upon expression of a K63R ubiquitin mutant protein (Supplementary Fig. 2). These results indicate that the second CARD of RIG-I is the primary site for Lys 63-linked ubiquitination.

To corroborate the ubiquitination of the 2CARD of RIG-I, six lysine residues were replaced with arginine (K→R) individually and in various combinations; these mutants were then tested for their ubiquitination level. The K172R mutation (alone or together with other mutations) caused near-complete loss of ubiquitination of the 2CARD of RIG-I (Fig. 1e and Supplementary Fig. 3a). In contrast, other K→R mutations had little or no effect on ubiquitination of the RIG-I 2CARD (Fig. 1e). As previously shown[10], wild-type GST–RIG-I(2CARD) potently induced IFN-β and NF-κB promoter activity (Fig. 1f and Supplementary Fig. 3b). GST–RIG-I(2CARD) mutants containing the K172R mutation alone or together with other mutations showed markedly reduced IFN-β and NF-κB promoter activation, consistent with their lack of ubiquitination; in contrast, other GST–RIG-I(2CARD) mutants induced IFN-β and NF-κB promoter activity as strongly as wild-type GST–RIG-I(2CARD) (Fig. 1f and Supplementary Fig. 3b). These results suggest that Lys 172 is the essential site for RIG-I 2CARD ubiquitination and signalling activity

[1]Department of Microbiology and Molecular Genetics and Tumor Virology Division, New England Primate Research Center, Harvard Medical School, 1 Pine Hill Drive, Southborough, Massachusetts 01772, USA. [2]Institute for Clinical and Molecular Virology, Friedrich-Alexander University Erlangen-Nuremberg, 91054 Erlangen, Germany. [3]Department of Microbiology, University of Ulsan College of Medicine, Seoul 138-736, South Korea. [4]Department of Geriatric Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8655, Japan. [5]Research Center for Genomic Medicine, Saitama Medical School, Saitama 350-124-2, Japan. [6]Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9148, USA. [7]Department of Host Defense, Japan Science and Technology Agency, Osaka 565-0871, Japan.

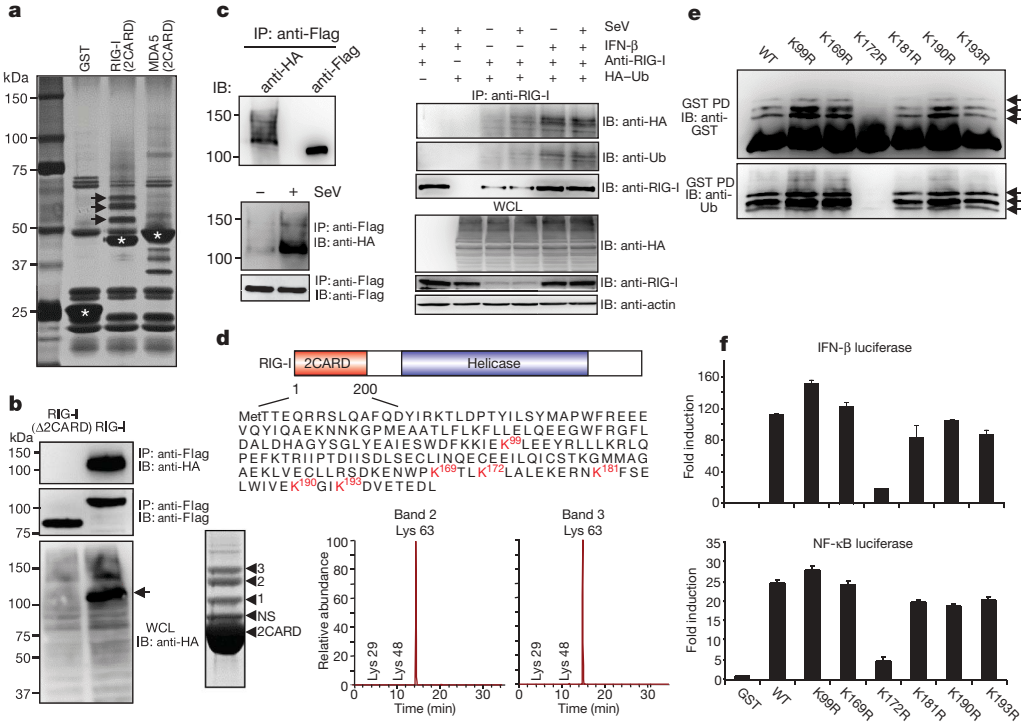


Figure 1 | The 2CARD of RIG-I undergoes robust ubiquitination. a, Silver-stained purified GST fusion complexes. Arrows, unique bands; asterisks, GST fusions. **b**, **c**, HEK293T cells transfected with Flag–RIG-I (**b**, and **c**, top-left) or Flag–RIG-I(Δ2CARD) (**b**) together with HA–ubiquitin were used for immunoprecipitation (IP) and immunoblotting (IB). WCL, whole cell lysate. **c**, Bottom-left panel: HEK293T cells transfected with Flag–RIG-I and HA–ubiquitin were mock-infected or infected with Sendai virus (SeV). Right: HEK293T cells transfected with HA–ubiquitin were treated (or not) with IFN-β and/or infected, as indicated, with Sendai virus before immunoprecipitation with anti-RIG-I antibody. **d**, The red lysine residues indicate the sites of ubiquitination. Bottom-left panel: Coomassie-blue-stained Flag–RIG-I(2CARD) complex. NS, nonspecific protein. Bottom-right panel: Lys 29/48/63-linked ubiquitination of RIG-I(2CARD)[21]. **e**, GST pull down (PD) of HEK293T cells transfected with GST–RIG-I(2CARD) or K→R mutants. Arrows indicate the ubiquitinated bands. WT, wild type. **f**, IFN-β and NF-κB promoter activity in GST–RIG-I(2CARD) or K→R mutant transfected cells. The results are expressed as means ± s.d. ($n = 3$).

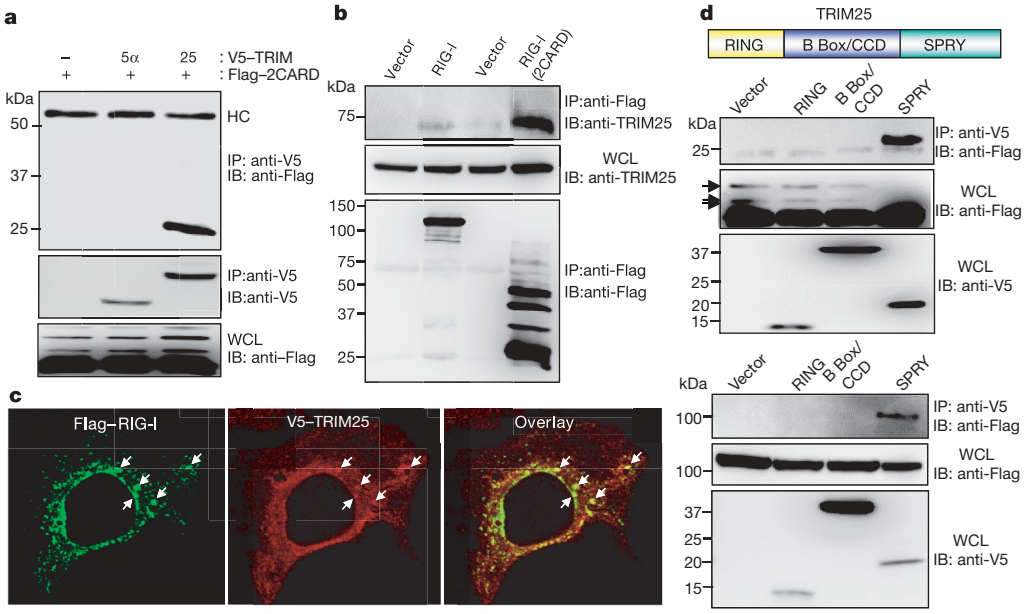


Figure 2 | Interaction between RIG-I and TRIM25. a, **b**, WCLs of HEK293T cells transfected with Flag–RIG-I(2CARD) and V5–TRIM25 or V5–TRIM5-α (**a**), or with Flag-RIG-I or Flag-RIG-I(2CARD) (**b**) were used for immunoprecipitation and immunoblotting, as indicated. **c**, Confocal images of Hela cells transiently transfected with Flag–RIG-I (green) and V5–TRIM25 (red). Arrows indicate representative co-localization between Flag–RIG-I and V5–TRIM25. Original magnification, ×100. **d**, WCLs of HEK293T cells transfected with Flag-tagged RIG-I(2CARD) (top) or full-length Flag–RIG-I (bottom) together with V5-tagged domains of TRIM25 were used for immunoprecipitation with V5 antibody, followed by immunoblotting with anti-Flag. WCLs were used for immunoblotting with anti-Flag and anti-V5 antibodies.

and that the extent of RIG-I 2CARD ubiquitination correlates strongly with its signal transduction activity.

Protein purification and mass spectrometry demonstrated that TRIM25 (also called oestrogen-responsive finger protein (EFP)[11]) is one of the proteins that associates with Flag–RIG-I(2CARD). TRIM25 has ubiquitin and ISG15 E3 ligase activity and downregulates 14-3-3σ through proteolysis for cell cycle regulation[12,13]. Co-immunoprecipitation revealed that RIG-I(2CARD) interacts with TRIM25 but not TRIM5-α, which has a similar structure to TRIM25 and functions as an intracellular inhibitor of retroviral replication[7] (Fig. 2a). Furthermore, interaction between Flag-tagged RIG-I or RIG-I(2CARD) and endogenous TRIM25 was readily detected in HEK293T cells (Fig. 2b). Confocal microscopy revealed that both RIG-I and TRIM25 exhibited punctate staining throughout the cytoplasm and that they co-localized extensively at cytoplasmic perinuclear bodies (Fig. 2c). As with other TRIM family members[7], TRIM25 contains a cluster of a RING-finger domain, a B box/coiled-coil domain (B Box/CCD) and a SPRY domain (Fig. 2d). Binding analysis revealed that the C-terminal SPRY domain of TRIM25 bound to both RIG-I and RIG-I(2CARD) as effectively as full-length TRIM25, whereas the RING-finger domain and B Box/CCD did not (Fig. 2d).

To test the role of TRIM25 in RIG-I ubiquitination, RIG-I or GST–RIG-I(2CARD) was co-expressed with wild-type TRIM25, E3 ligase-defective TRIM25(ΔRING) or TRIM5-α. TRIM25 expression markedly increased the ubiquitination levels of exogenous RIG-I and GST–RIG-I(2CARD), as well as endogenous RIG-I, but neither TRIM25(ΔRING) nor TRIM5-α had any effect (Fig. 3a, b; see also Supplementary Fig. 4a). In contrast, TRIM25 expression did not induce the ubiquitination of GST–MDA5(2CARD) (Supplementary Fig. 4b). TRIM25 depletion *in vivo* by a TRIM25-specific small hairpin RNA (shRNA)[13] significantly reduced the ubiquitination level of GST–RIG-I(2CARD) and RIG-I in a dose-dependent manner (Fig. 3c and Supplementary Fig. 5a), but a nonspecific scrambled-sequence shRNA had no effect on GST–RIG-I(2CARD) ubiquitination (Supplementary Fig. 5b). Finally, an *in vitro* ubiquitination assay showed that TRIM25 effectively delivered the ubiquitin moieties to maltose-binding protein (MBP)–T7-tagged RIG-I(2CARD), but not MBP–T7 alone or MBP–T7–RIG-I(2CARD(170stop)) (Fig. 3d and Supplementary Fig. 6a). Consistent with its ubiquitination level, RIG-I-mediated induction of IFN-β or NF-κB promoter activity considerably increased on TRIM25 expression in a dose-dependent manner (Fig. 3e and Supplementary Fig. 6b). Notably, expression of the TRIM25(SPRY) mutant, which was sufficient to bind to RIG-I, markedly suppressed GST–RIG-I(2CARD) ubiquitination in a dose-dependent manner (Fig. 3f) as well as endogenous RIG-I ubiquitination (Supplementary Fig. 4a). Furthermore, expression of the TRIM25(SPRY) mutant considerably decreased the RIG-I 2CARD-mediated activation of IFN-β or NF-κB promoter activity in a dose-dependent manner (Supplementary Fig. 7). This suggests that TRIM25-mediated ubiquitination has an important role in RIG-I signalling activity.

Unlike the GST–RIG-I(2CARD) K172R mutant, which showed an almost complete loss of ubiquitination and IFN-β and NF-κB promoter activation, GST–RIG-I(2CARD) K172only—containing five K→R substitutions but leaving K172 intact—demonstrated highly induced IFN-β and NF-κB promoter activity (Fig. 4a, b). Furthermore, the GST–RIG-I(2CARD) K172only mutant underwent robust ubiquitination (albeit lower than that of wild-type GST–RIG-I(2CARD)) on TRIM25 expression, whereas GST–RIG-I(2CARD) K172R was minimally ubiquitinated (Fig. 4c). However, despite a significant reduction in its level of ubiquitination, GST–RIG-I(2CARD) K172R interacted with TRIM25 as efficiently as wild-type GST–RIG-I(2CARD) and the K172only mutant (Fig. 4c). As seen with RIG-I(2CARD), full-length RIG-I K172only but not RIG-I K172R demonstrated ubiquitination at the same level as RIG-I wild type (Supplementary Fig. 8). Finally, correlated with their ubiquitination levels, expression of wild-type RIG-I and mutant RIG-I K172only in *RIG-I*$^{-/-}$ mouse embryonic fibroblasts (MEFs) induced IFN-β production on Sendai virus infection, whereas expression of mutant RIG-I K172R showed no effect on IFN-β production (Supplementary Fig. 9).

The 2CARD of RIG-I has been shown to bind to the MAVS CARD to elicit downstream signal transduction[14–17]. GST pull-down analysis showed that wild-type GST–RIG-I(2CARD) and mutant GST–RIG-I(2CARD) K172only efficiently interacted with the Flag-tagged CARD proline-rich domain of MAVS (Flag–MAVS(CARD-PRD)), whereas GST–RIG-I(2CARD) K172R and GST–RIG-I(2CARD) K99,169,172, 181,190,193R mutants poorly bound to Flag–MAVS(CARD-PRD) (Fig. 4d), indicating that Lys 172 is critical for TRIM25-mediated ubiquitination, RIG-I signalling and MAVS interaction, but not for TRIM25 binding (Fig. 4c).

Wild-type, *Trim25*$^{+/-}$ and *Trim25*$^{-/-}$ MEFs[18] were used to test the direct contribution of TRIM25 to RIG-I-mediated IFN-β expression. IFN-β promoter activity was very low in *Trim25*$^{-/-}$ MEFs and was reduced in *Trim25*$^{+/-}$ MEFs compared with wild-type MEFs

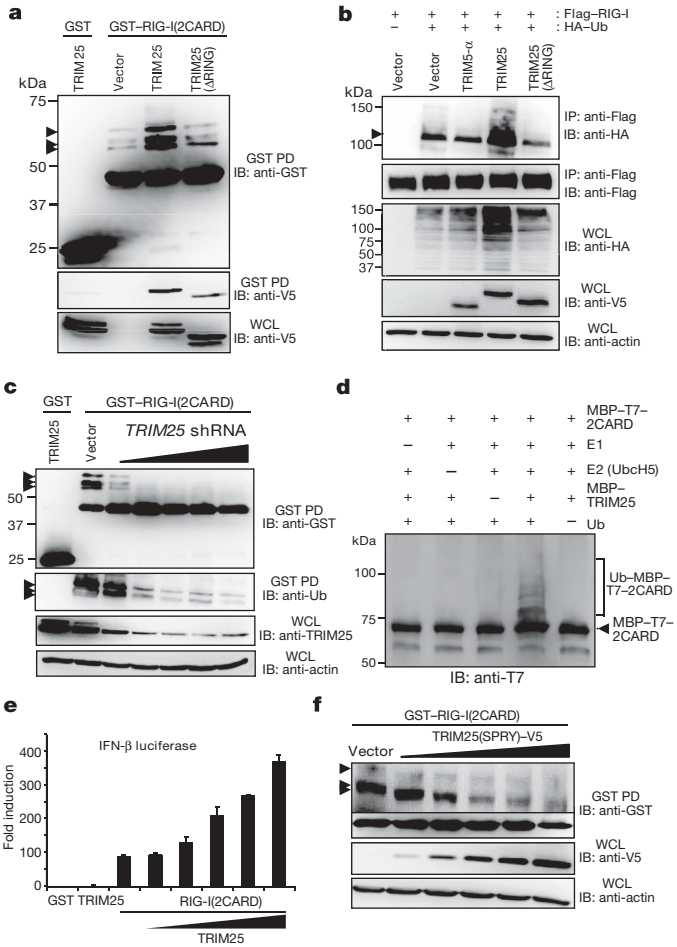


Figure 3 | TRIM25 is a primary E3 ubiquitin ligase of RIG-I. HEK293T cells transfected with GST or GST–RIG-I(2CARD) (**a**) or Flag–RIG-I and HA–ubiquitin (**b**) together with vector, TRIM25, TRIM25(ΔRING) or TRIM5-α were used for GST pull down (PD) (**a**) or immunoprecipitation with anti-Flag antibody (**b**). **c**, HEK293T cells transfected with GST or GST–RIG-I(2CARD) together with pSUPER.retro.puro or *TRIM25*-shRNA-specific pSUPER.retro.puro[13] were used for GST pull down. Arrows indicate the ubiquitinated GST–RIG-I(2CARD) and Flag–RIG-I. **d**, *In vitro* ubiquitination was detected by anti-T7 immunoblotting. **e**, IFN-β luciferase activity in HEK293T cells transfected with GST–RIG-I(2CARD) and TRIM25. The results are expressed as means ± s.d. ($n = 3$). **f**, HEK293T cells transfected with GST–RIG-I(2CARD) and V5–TRIM25(SPRY) were used for GST pull down. Arrows indicate ubiquitinated GST–RIG-I(2CARD).

(Supplementary Fig. 10a). Consistent with IFN-β promoter activation, virus-induced IFN-β production was virtually undetectable in *Trim25*$^{-/-}$ MEFs, whereas it was considerably high in wild-type MEFs (Fig. 4e). *Trim25*$^{+/-}$ MEFs showed a slightly reduced level of IFN-β production compared with wild-type MEFs (Fig. 4e). On vesicular stomatitis virus (VSV)–enhanced green fluorescent protein (eGFP) infection at various multiplicity of infections (MOIs), *Trim25*$^{-/-}$ MEFs showed remarkably increased levels of VSV–eGFP-positive cells (Fig. 4g and Supplementary Fig. 10b) and increased VSV yields (over 100-fold) (Fig. 4f) compared with wild-type and *Trim25*$^{+/-}$ MEFs. Similarly, *Trim25*$^{-/-}$ MEFs showed a considerable increase in the level of Newcastle disease virus (NDV)–GFP infection (Supplementary Fig. 10c). Finally, TRIM25 expression significantly suppressed VSV–eGFP replication in HEK293T cells, whereas expression of the TRIM25(SPRY) mutant detectably increased VSV–eGFP replication (Fig. 4h). Collectively, these results indicate that TRIM25 is critical for cytosolic RIG-I signal transduction that mediates the induction of the IFN response on viral infection.

Ubiquitination is a versatile post-translational modification involved in various cellular functions[19]. Our study indicates that TRIM25 E3 ubiquitin ligase induces the Lys 63-linked ubiquitination of RIG-I; that Lys 172 is the critical site for TRIM25-mediated ubiquitination; and that, as seen with the ubiquitin-dependent interaction between RIP and NEMO[20], the TRIM25-mediated ubiquitination of RIG-I may facilitate its interaction with MAVS, which ultimately leads to downstream signal transduction. Thus, the interconnection between the RIG-I cytosolic viral RNA receptor and a member of the TRIM family represents a new class of antiviral regulatory pathway involved in innate immunity.

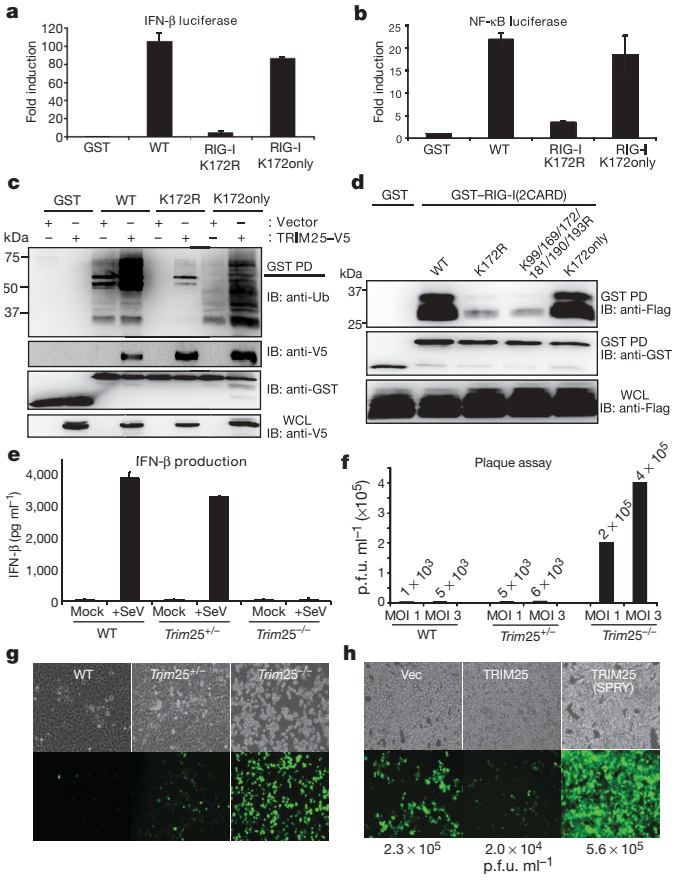


Figure 4 | Role of TRIM25-mediated ubiquitination in RIG-I antiviral activity. IFN-β (**a**) and NF-κB (**b**) promoter activity in GST–RIG-I(2CARD) or mutant transfected cells. The results are expressed as means ± s.d. ($n = 3$). **c**, HEK293T cells transfected with GST–RIG-I(2CARD), GST–RIG-I(2CARD) K172R, or GST–RIG-I(2CARD) K172only together with vector or TRIM25 were used for GST pull down. **d**, HEK293T cells transfected with GST–RIG-I(2CARD) or the indicated mutants together with Flag–MAVS(CARD-PRD) were used for GST pull down. **e**, IFN-β production in wild-type, *Trim25*$^{+/-}$ and *Trim25*$^{-/-}$ MEFs upon Sendai virus infection. The results are expressed as means ± s.d. ($n = 3$). **f–h**, VSV–eGFP replication in wild-type, *Trim25*$^{+/-}$ and *Trim25*$^{-/-}$ MEFs (**f** and **g**) or in vector-, TRIM25-, or TRIM25(SPRY)-expressing HEK293T cells (**h**) was determined by plaque assay or visualized by fluorescence microscopy.

METHODS SUMMARY

RNA interference for TRIM25. The mammalian expression vector pSUPER.retro.puro (OligoEngine), encoding shRNAs for *TRIM25* sequence, was provided by D.-E. Zhang. Details of the shRNA sequence and transfection method have been described[13].

Viruses. NDV–GFP and VSV–eGFP were provided by A. Garcia-Sastre and S. Whelan, respectively.

Measurement of IFN-β production. Cell culture supernatants were collected and analysed for IFN-β production using enzyme-linked immunosorbent assays (PBL Biomedical Laboratories).

***In vitro* ubiquitination assay.** Purified MBP–T7–RIG-I(2CARD) (20 μg ml^{-1}) and MBP–TRIM25 (20 μg ml^{-1}) derived from *Escherichia coli* were incubated in a reaction buffer (50 mM Tris-HCl, 2 mM dithiothreitol, 5 mM $MgCl_2$ and 4 mM ATP) with ubiquitin (50 μg ml^{-1}; Sigma), human recombinant E1 (1.6 μg ml^{-1}; BIOMOL) and human recombinant UbcH5a (20 μg ml^{-1}; BIOMOL) at 32 °C for 2 h and subjected to immunoblotting with anti-T7 antibody (Novagen).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 February; accepted 8 March 2007.
Published online 28 March 2007.

1. Honda, K., Takaoka, A. & Taniguchi, T. Type I interferon gene induction by the interferon regulatory factor family of transcription factors. *Immunity* 25, 349–360 (2006); erratum **25**, 849 (2006).
2. Hornung, V. *et al.* 5′-Triphosphate RNA is the ligand for RIG-I. *Science* **314**, 994–997 (2006).
3. Meylan, E. & Tschopp, J. Toll-like receptors and RNA helicases: two parallel ways to trigger antiviral responses. *Mol. Cell* **22**, 561–569 (2006).
4. Pichlmair, A. *et al.* RIG-I-mediated antiviral responses to single-stranded RNA bearing 5′-phosphates. *Science* **314**, 997–1001 (2006).
5. Stetson, D. B. & Medzhitov, R. Antiviral defense: interferons and beyond. *J. Exp. Med.* **203**, 1837–1841 (2006).
6. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
7. Nisole, S., Stoye, J. P. & Saib, A. TRIM family proteins: retroviral restriction and antiviral defence. *Nature Rev. Microbiol.* **3**, 799–808 (2005).
8. Johnson, C. L. & Gale, M. Jr. CARD games between virus and host get a new player. *Trends Immunol.* **27**, 1–4 (2006).
9. Meylan, E., Tschopp, J. & Karin, M. Intracellular pattern recognition receptors in the host response. *Nature* **442**, 39–44 (2006).
10. Kato, H. *et al.* Cell type-specific involvement of RIG-I in antiviral response. *Immunity* **23**, 19–28 (2005).
11. Orimo, A., Inoue, S., Ikeda, K., Noji, S. & Muramatsu, M. Molecular cloning, structure, and expression of mouse estrogen-responsive finger protein Efp. Co-localization with estrogen receptor mRNA in target organs. *J. Biol. Chem.* **270**, 24406–24413 (1995).
12. Urano, T. *et al.* Efp targets 14-3-3σ for proteolysis and promotes breast tumour growth. *Nature* **417**, 871–875 (2002).
13. Zou, W. & Zhang, D. E. The interferon-inducible ubiquitin-protein isopeptide ligase (E3) EFP also functions as an ISG15 E3 ligase. *J. Biol. Chem.* **281**, 3989–3994 (2006).
14. Meylan, E. *et al.* Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).
15. Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-κB and IRF 3. *Cell* **122**, 669–682 (2005).
16. Xu, L. G. *et al.* VISA is an adapter protein required for virus-triggered IFN-β signaling. *Mol. Cell* **19**, 727–740 (2005).
17. Kawai, T. *et al.* IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunol.* **6**, 981–988 (2005).
18. Orimo, A. *et al.* Underdeveloped uterus and reduced estrogen responsiveness in mice with disruption of the estrogen-responsive finger protein gene, which is a direct target of estrogen receptor α. *Proc. Natl Acad. Sci. USA* **96**, 12027–12032 (1999).

19. Haglund, K. & Dikic, I. Ubiquitylation and cell signaling. *EMBO J.* **24,** 3353–3359 (2005).
20. Ea, C. K., Deng, L., Xia, Z. P., Pineda, G. & Chen, Z. J. Activation of IKK by TNFα requires site-specific ubiquitination of RIP1 and polyubiquitin binding by NEMO. *Mol. Cell* **22,** 245–257 (2006).
21. Kirkpatrick, D. S., Denison, C. & Gygi, S. P. Weighing in on ubiquitin: the expanding role of mass-spectrometry-based proteomics. *Nature Cell Biol.* **7,** 750–757 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by US Public Health Service grants (J.U.J.), the exchange programme between Harvard Medical School and the graduate training programme 1071 at the Friedrich-Alexander University Erlangen-Nuremberg, Germany (M.U.G.), and a Korea Research Foundation Grant (C.-H.J.). We thank A. Garcia-Sastre, D.-E. Zhang and S. Whelan for providing reagents, and R. Tomaino and J. Nagel for mass spectrometry. We also thank all members of the Tumor Virology Division, New England Primate Research Center, for discussions.

Author Contributions M.U.G. performed all aspects of this study. Y.C.S., C.-H.J. and C.L. assisted in experimental design and in collecting the data. T.U. and S.I. performed the *in vitro* ubiquitination assay and generated *Trim25*$^{-/-}$ MEFs. L.S. and Z.C. generated the MAVS construct and RIG-I antibody. T.O. and S.A. generated the RIG-I construct and *RIG-I*$^{-/-}$ MEFs. M.U.G. and J.U.J. organized this study and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.U.J. (jae_jung@hms.harvard.edu).

METHODS

Cell culture. HEK293T, MEF and Hela cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine and 1% penicillin-streptomycin (Gibco-BRL). Transient transfections were performed with FuGENE 6 (Roche), lipofectamine 2000 (Invitrogen), or calcium phosphate (Clontech) following the manufacturer's instructions. Wild-type, *Trim25*$^{+/-}$ and *Trim25*$^{-/-}$ MEFs were immortalized with LXSN-E6/E7 retroviral vector containing human papilloma virus 16 E6 and E7 oncogenes using a standard protocol of selection with 200 μg ml^{-1} of neomycin. *RIG-I*$^{-/-}$ MEFs were infected with pBabe-puro vector, pBabe-puro-RIG-I wild-type, pBabe-puro-RIG-I K172R, or pBabe-puro-RIG-I K172only retrovirus, followed by selection with 1 μg ml^{-1} of puromycin.

Plasmid construction. All constructs for transient and stable expression in mammalian cells were derived from the pEBG GST fusion vector and the pEF-IRES-Puro expression vector. DNA fragments corresponding to the coding sequence of the *RIG-I* and *TRIM25* genes were amplified from template DNA by polymerase chain reaction (PCR) and subcloned into plasmid pEBG between restriction sites *Kpn*I and *Not*I or pEF-IRES-puro between *Afl*II and *Not*I for selection of stable transfectants. V5-tagged TRIM25 and Flag-tagged RIG-I were expressed from a modified pIRES-puro encoding a C-terminal V5 tag and Flag tag, respectively. RIG-I mutants were generated by PCR using site-directed mutagenesis. All constructs were sequenced using an ABI PRISM 377 automatic DNA sequencer to verify 100% agreement with the original sequence.

***In vivo* GST pull down, protein purification and mass spectrometry.** At 48 h after transfection with vectors expressing GST, GST–RIG-I(2CARD) or GST–MDA5(2CARD) fusions, HEK293T cells were collected and lysed with NP40 buffer (50 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% (v/v) NP40) supplemented with a complete protease inhibitor cocktail (Roche). Post-centrifuged supernatants were pre-cleared with protein A/G beads at 4 °C for 2 h. Pre-cleared lysates were mixed with a 50% slurry of glutathione-conjugated Sepharose beads (Amersham Biosciences), and the binding reaction was incubated for 4 h at 4 °C. Precipitates were washed extensively with lysis buffer. Proteins bound to glutathione beads were eluted and separated on a NuPAGE 4–12% Bis-Tris gradient gel (Invitrogen). After Coomassie or silver staining (Invitrogen), specific protein bands were excised and analysed by ion-trap mass spectrometry at the Harvard Taplin Biological Mass Spectrometry facility, and amino acid sequences were determined by tandem mass spectrometry and database searches.

Immunoblot analysis and immunoprecipitation assay. For immunoblotting, polypeptides were resolved by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and transferred to a PVDF membrane (Bio-Rad). Immunodetection was achieved with anti-V5 (1:5,000) (Invitrogen), anti-Flag (1:5,000) (Sigma), anti-HA (1:5,000), anti-GST (1:10,000) (Sigma), anti-actin (1:10,000) (Abcam), or anti-TRIM25 (1:2,000) (BD Bioscience) antibodies. The proteins were visualized by a chemiluminescence reagent (Pierce) and detected by a Fuji Phosphor Imager.

For immunoprecipitation, cells were collected after 48 h and then lysed in NP40 buffer supplemented with a complete protease inhibitor cocktail (Roche). After pre-clearing with protein A/G agarose beads for 2 h at 4 °C, whole-cell lysates were used for immunoprecipitation with the indicated antibodies. Generally, 1–2 μg of commercial antibody was added to 1 ml of cell lysate, which was incubated at 4 °C for 4–12 h. After addition of protein A/G agarose beads, the incubation was continued for 2 h. Immunoprecipitates were extensively washed with lysis buffer and eluted with SDS loading buffer by boiling for 5 min.

Confocal immunofluorescence microscopy. Eighteen to twenty-four hours after transfection, cells were fixed with 4% paraformaldehyde for 15 min, permeabilized with 0.2% (v/v) Triton X-100 for 15 min, blocked with 10% goat serum in PBS for 1 h and reacted with diluted primary antibody in 1% goat serum for up to 2 h at room temperature. After incubation, cells were washed extensively with PBS, incubated with the appropriate secondary antibody diluted in 1% goat serum for 1 h at room temperature, and washed three times with PBS. Confocal microscopy was performed using a Leica TCS SP laser-scanning microscope (Leica Microsystems) fitted with a ×100 Leica objective (PL APO, 1.4NA) and Leica imaging software. Images were collected at 512 × 512-pixel resolution. The stained cells were optically sectioned in the *z* axis, and the images in the different channels (photo multiplier tubes) were collected simultaneously. The step size in the *z* axis varied from 0.2 to 0.5 μm to obtain 16 slices per imaged file. The images were transferred to a Macintosh G4 computer (Apple Computer), and Photoshop (Adobe) was used to render the images.

LETTERS

Ubiquitination by the anaphase-promoting complex drives spindle checkpoint inactivation

S. K. Reddy[1,2]*, M. Rape[1]*†, W. A. Margansky[1] & M. W. Kirschner[1]

Eukaryotic cells rely on a surveillance mechanism known as the spindle checkpoint to ensure accurate chromosome segregation. The spindle checkpoint prevents sister chromatids from separating until all kinetochores achieve bipolar attachments to the mitotic spindle[1–3]. Checkpoint proteins tightly inhibit the anaphase-promoting complex (APC), a ubiquitin ligase required for chromosome segregation and progression to anaphase. Unattached kinetochores promote the binding of checkpoint proteins Mad2 and BubR1 to the APC-activator Cdc20, rendering it unable to activate APC. Once all kinetochores are properly attached, however, cells inactivate the checkpoint within minutes, allowing for the rapid and synchronous segregation of chromosomes[4]. How cells switch from strong APC inhibition before kinetochore attachment to rapid APC activation once attachment is complete remains a mystery. Here we show that checkpoint inactivation is an energy-consuming process involving APC-dependent multi-ubiquitination. Multi-ubiquitination by APC leads to the dissociation of Mad2 and BubR1 from Cdc20, a process that is reversed by a Cdc20-directed de-ubiquitinating enzyme[5]. The mutual regulation between checkpoint proteins and APC leaves the cell poised for rapid checkpoint inactivation and ensures that chromosome segregation promptly follows the completion of kinetochore attachment. In addition, our results suggest a mechanistic basis for how cancer cells can have a compromised spindle checkpoint without corresponding mutations in checkpoint genes[6].

Cells exposed to spindle poisons such as nocodazole or taxol arrest in prometaphase owing to activation of the spindle checkpoint by unattached kinetochores[6]. In previous studies on APC, we found that cells overexpressing the APC-specific ubiquitin-conjugating enzyme (E2), UbcH10, showed a compromised arrest in mitosis in response to nocodazole[7]. To characterize this phenotype more precisely, we transfected HeLa cells with wild-type UbcH10, a catalytically dead active-site mutant—UbcH10^{C114S}—or empty vector, and subsequently challenged them with nocodazole. Whereas cells transfected with empty vector or UbcH10^{C114S} were efficiently arrested in mitosis, nearly 70% of cells overexpressing wild-type UbcH10 were in interphase (Fig. 1a). Cells with a defective spindle checkpoint can exit mitosis even in the presence of spindle poisons, generating multinucleated cells with decondensed interphase-like chromatin. Multinucleated cells were rarely observed among taxol-treated HeLa cells overexpressing UbcH10^{C114S} or empty vector (Fig. 1b). Under the same conditions, nearly half of the interphase cells overexpressing UbcH10 were multinucleated. These data strongly suggest that cells expressing high levels of UbcH10 can enter mitosis, but are unable to maintain spindle checkpoint activity.

To biochemically characterize the effect of UbcH10, we prepared checkpoint extracts (CP-extracts) from HeLa cells arrested in mitosis by nocodazole treatment. As reported previously[8], CP-extracts failed to degrade substrates of APC (Supplementary Fig. 1a). Addition of UbcH10 to CP-extracts, however, inactivated the spindle checkpoint, as indicated by the degradation of substrates (geminin, securin and cyclin B1) of the APC^{Cdc20} complex (Fig. 1c and Supplementary Fig. 1a). This degradation was APC-dependent because it could be reversed by the addition of the APC inhibitor Emi1, or by additional Mad2 (Supplementary Fig. 1a, b). APC purified from CP-extracts (CP-APC) was also inactivated by the spindle checkpoint. CP-APC was unable to ubiquitinate substrates in the absence of exogenous E2 but addition of either UbcH10 or the other APC-interacting E2, UbcH5α, to CP-APC promoted the ubiquitination of APC^{Cdc20} substrates (Supplementary Fig. 1c, d). Together, these results demonstrate

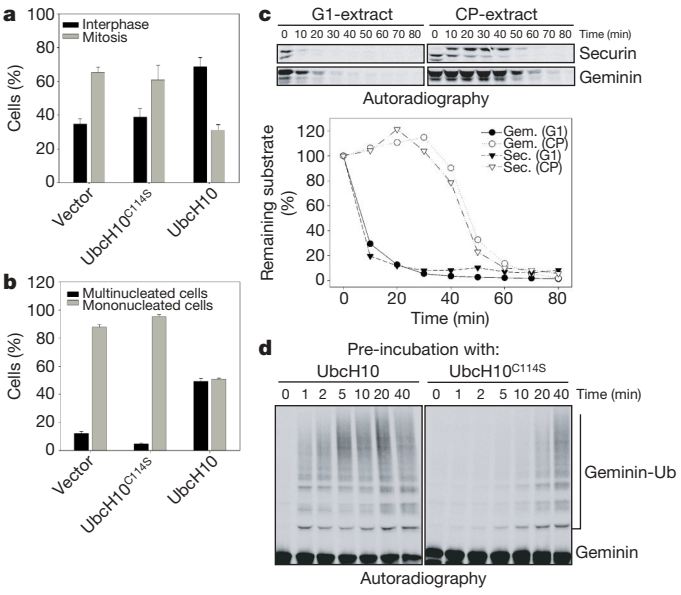


Figure 1 | UbcH10 activates CP-APC for substrate ubiquitination and degradation. a, UbcH10 overrides the spindle checkpoint in nocodazole-arrested cells. Cells were transfected with indicated plasmids and treated with nocodazole 4 h later. After 16 h the mitotic index was determined in transfected cells. Error bars, s.e.m. ($n = 3$). **b**, Overexpression of UbcH10, but not of inactive UbcH10^{C114S}, causes multinucleation. Error bars, s.e.m. ($n = 3$). **c**, APC-substrates are degraded rapidly in G1-extracts but display a lag before degradation in CP-extracts. G1-extracts or CP-extracts were incubated with UbcH10, and the abundance of ^{35}S-labelled substrates was analysed by autoradiography. Gem., geminin; sec., securin. **d**, Pre-incubation with UbcH10 but not with UbcH10^{C114S} eliminates the lag in substrate ubiquitination by purified CP-APC.

[1]Department of Systems Biology, Harvard Medical School, and [2]Harvard–MIT Division of Health Sciences and Technology, Boston, Massachusetts 02115, USA. †Present address: Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720-3202, USA.
*These authors contributed equally to this work.

that UbcH10 can directly override inhibition of APC by the spindle checkpoint *in vivo*, in extracts and in the purified system.

APC substrates such as securin and geminin are rapidly degraded in extracts prepared from G1 cells (G1-extracts)[8] (Fig. 1c). In contrast, these substrates demonstrated biphasic kinetics in UbcH10-treated CP-extracts, with little substrate degradation in the first 40 minutes of reaction followed by rapid degradation thereafter. We observed a similar lag phase for substrate ubiquitination by purified CP-APC (Supplementary Fig. 1e). These results indicated that UbcH10 may have two roles on CP-APC, an initial role in checkpoint inactivation, which occurs during the lag phase, and a subsequent role in ubiquitin transfer to substrates. To test this, we pre-incubated CP-extracts with either wild-type UbcH10 or inactive UbcH10^{C114S} to allow any inactivation of the checkpoint to occur. We then purified APC from these extracts, washed away the UbcH10 variants, and added fresh, wild-type UbcH10 along with ubiquitination components to observe the kinetics of substrate ubiquitination. Markedly, pre-incubation of CP-extracts with wild-type UbcH10 but not UbcH10^{C114S} completely eliminated the lag in substrate ubiquitination (Fig. 1d). These results strongly suggest that UbcH10 has a direct role in overriding checkpoint inhibition of APC and that this requires UbcH10's catalytic activity.

Because the stability of Mad2, BubR1, Bub3 and Mad1 in CP-extracts was unaffected by UbcH10 addition, the initial inactivation of the checkpoint is unlikely to involve the destruction of checkpoint proteins (Supplementary Fig. 2a). Previous studies had shown that association of Mad2 and BubR1 with APC is necessary for proper checkpoint function[9–11]. Therefore, we examined whether UbcH10 affected this association. In the absence of UbcH10, the interaction of Mad2 and BubR1 with CP-APC was stable, with a half-life of several hours (Supplementary Fig. 2b). Furthermore, in CP-extracts incubated with UbcH10^{C114S}, Mad2 and BubR1 co-migrated with APC as analysed by sucrose gradient centrifugation (Fig. 2a). In contrast, incubation of CP-extracts with wild-type UbcH10 led to the complete loss of Mad2 and BubR1 from APC (Fig. 2a and Supplementary Fig. 2c). Similarly, purified CP-APC incubated with UbcH10 and ubiquitination components E1 and ubiquitin, showed complete dissociation of Mad2 from APC as well as dissociation of BubR1 and Cdc20 (Fig. 2b). Finally, UbcH10 promoted the dissociation of checkpoint proteins *in vivo*. In cells depleted of UbcH10 by RNA interference (RNAi), dissociation of Mad2 from Cdc20 was delayed after release from nocodazole (Supplementary Fig. 2d). This corresponds with a delay in completing the metaphase to anaphase transition as observed by live-cell imaging of UbcH10-depleted cells that had not been exposed to spindle toxins (Fig. 2c and Supplementary Fig. 3). Taken together, these results demonstrate that catalytically active UbcH10 can promote the release of checkpoint proteins from APC.

We next explored the detailed requirements for UbcH10 activity during checkpoint inactivation. When CP-extracts were incubated with UbcH10 and wild-type ubiquitin, checkpoint proteins dissociated from APC as expected (Fig. 3a). In contrast, incubation of CP-extracts with UbcH10 and methylubiquitin or lysine-free ubiquitin failed to promote dissociation of checkpoint proteins from APC (Fig. 3a and Supplementary Fig. 4b). Both methylubiquitin and lysine-free ubiquitin are competent to mono-ubiquitinate substrates but not to form multi-ubiquitin chains, suggesting that multi-ubiquitination by APC is required for checkpoint inactivation. Multi-ubiquitination usually marks substrates for proteasomal proteolysis; however, whether extracts were treated with UbcH10 alone or UbcH10 in the presence of the proteasome inhibitor MG132, Mad2 and BubR1 still dissociated from APC (Fig. 3b). Furthermore, MG132 treatment prevented the partial proteolysis of Cdc20 observed on UbcH10 addition, but did not interfere with checkpoint complex disassembly. Thus, proteasomal proteolysis does not seem to be essential for the disassembly of checkpoint complexes, although we cannot rule out a role for Cdc20 degradation at steps subsequent to Mad2 dissociation. We conclude that release of checkpoint proteins requires multi-ubiquitination of one or more components of the APC–Cdc20–checkpoint protein complex.

While examining Mad2 and BubR1 release from purified CP-APC, we noticed that co-purifying Cdc20 showed a characteristic ladder of multi-ubiquitination that was dependent on UbcH10 (Supplementary Fig. 4c). We also detected ubiquitinated Cdc20 *in vivo* from nocodazole-arrested cells. This ubiquitination was maintained even after short interfering (si)RNA-mediated depletion of Cdh1, the APC activator responsible for Cdc20 ubiquitination in late mitosis (Supplementary Fig. 4d). These results suggest that CP-APC directly ubiquitinates Cdc20 in early mitosis. Importantly, multi-ubiquitination of Cdc20 in UbcH10-treated CP-extracts coincided with lower amounts of co-purifying Mad2 (Fig. 3c). Furthermore, in reactions with purified CP-APC, Cdc20 ubiquitination preceded the onset of substrate ubiquitination (Fig. 3d and Supplementary Fig. 1e). Thus, Cdc20 ubiquitination is an early event that may contribute to disassembly of checkpoint complexes by CP-APC. In a parallel study, a Cdc20-directed de-ubiquitinating enzyme, USP44, was shown to counteract UbcH10-mediated checkpoint inactivation both *in vitro* and *in vivo*[5]. Purified USP44 directly promotes de-ubiquitination of Cdc20 (ref. 5). Additionally, when USP44 is depleted from cells, Cdc20 ubiquitination is increased and the amount of Mad2 associated with Cdc20 is markedly reduced. Co-depletion of UbcH10 reverses both phenotypes[5]. Taken together, these results strongly suggest that APC-dependent ubiquitination of Cdc20 decreases its association with Mad2 and thus frees APC to ubiquitinate substrates.

Mad2 and Cdc20 are intimately associated in an energetically stable complex when the checkpoint is active[12–14]. Ubiquitination

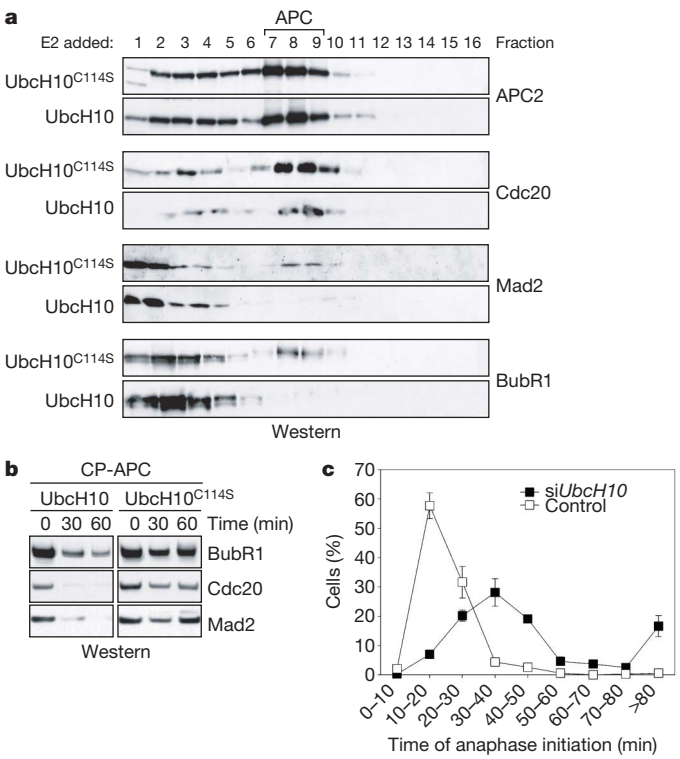


Figure 2 | UbcH10 catalyses the dissociation of checkpoint components from APC. a, Checkpoint proteins dissociate from APC in the presence of UbcH10. CP-extracts were incubated with UbcH10 or UbcH10^{C114S} along with ubiquitination components and centrifuged through sucrose gradients. Indicated proteins were detected by western analysis. **b**, UbcH10 promotes dissociation of checkpoint proteins from purified CP-APC. **c**, Depletion of UbcH10 delays the onset of anaphase. HeLa cells expressing GFP–histone-2B were treated with control or with *UBCH10* siRNA and monitored by fluorescence microscopy. Error bars, s.e.m. ($n = 3$).

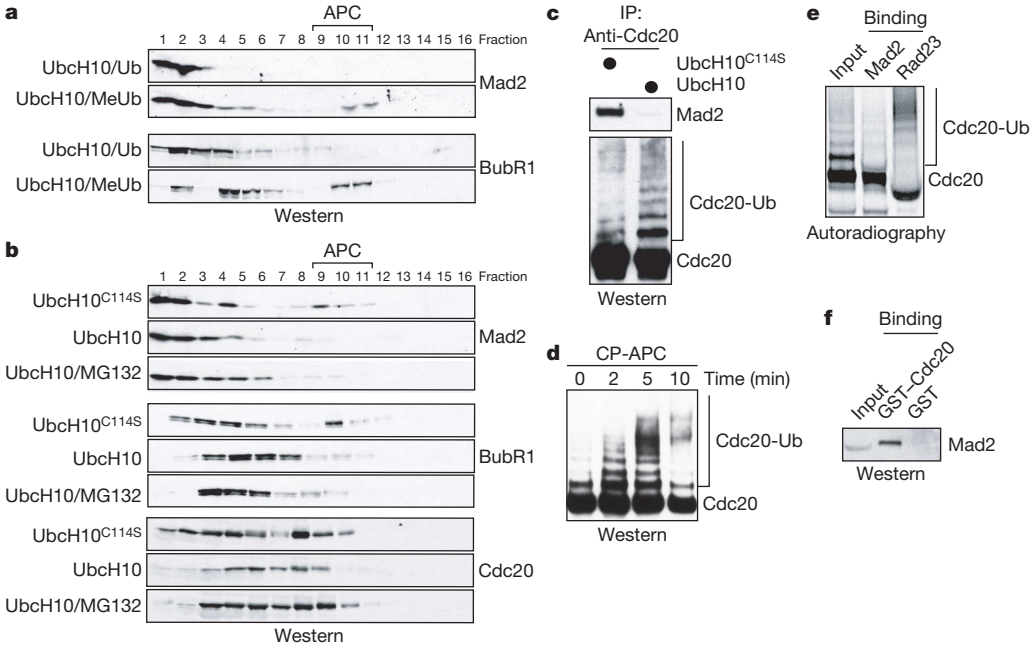


Figure 3 | Disassembly of checkpoint complexes requires APC-dependent multi-ubiquitination. a. UbcH10-dependent dissociation of checkpoint proteins requires multi-ubiquitination. CP-extracts were incubated with ubiquitin (Ub) or methylubiquitin before sucrose gradient centrifugation. **b**, Dissociation of checkpoint proteins does not require proteolysis. MG132 is a proteasome inhibitor. **c**, UbcH10 but not UbcH10^{C114S} promotes Cdc20 ubiquitination and dissociation of Mad2 from Cdc20. CP-extracts were incubated with UbcH10 or UbcH10^{C114S}. Cdc20 ubiquitination and the amount of co-purifying Mad2 were determined by western analysis. **d**, Cdc20 ubiquitination occurs during the lag-phase of substrate ubiquitination seen with CP-APC. Purified CP-APC was incubated with UbcH10 and ubiquitination components to determine kinetics of ubiquitination of co-purifying Cdc20. **e**, Ubiquitinated Cdc20 cannot bind Mad2. ^{35}S-labelled Cdc20 was exchanged into CP-APC and ubiquitinated *in vitro*. Cdc20 was then incubated with recombinant Mad2- or Rad23-conjugated Ni-NTA-agarose beads. Bound Cdc20 was analysed by autoradiography. **f**, Mad2 released from CP-APC seems to undergo a conformational change. Mad2 was released from CP-APC by UbcH10 addition and subsequently incubated with glutathione sepharose beads pre-conjugated with GST–Cdc20 peptide or GST alone. Bound Mad2 was detected by western analysis.

of Cdc20 might be expected to interfere with Mad2 binding by eliciting a conformational change in Cdc20 or by direct steric occlusion of the Mad2 binding surface. Consistent with this hypothesis, we found that purified Mad2 failed to bind ubiquitinated Cdc20, though it efficiently bound unmodified Cdc20 (Fig. 3e). Additionally, Mad2 liberated from CP-APC by the addition of UbcH10 was efficiently bound by a GST-tagged peptide of Cdc20 (Fig. 3f). Because Mad2 in CP-APC is locked in a closed or N2′ conformation, and because this confomer is incapable of binding the Cdc20 peptide[15,16], APC-dependent multi-ubiquitination seems to induce a substantial conformational change in Mad2. Together, these results suggest that APC-dependent multi-ubiquitination leads to the release of Mad2 from Cdc20 and blocks rebinding of Mad2 to ubiquitinated Cdc20.

In earlier work p31^{comet}, a protein that binds Cdc20-bound Mad2, was implicated in spindle checkpoint inactivation[17,18]. We found that, as with UbcH10, addition of p31^{comet} to CP-extracts promoted the degradation of APC substrates (Fig. 4a). Addition of both UbcH10 and p31^{comet} markedly accelerated the rate of substrate degradation beyond that observed with either protein alone. Furthermore, during substrate ubiquitination by CP-APC, p31^{comet} lowered the threshold for UbcH10 action to the physiological concentrations of UbcH10 in mitosis (100 nM; Fig. 4b). Thus, p31^{comet} and UbcH10 work in concert during checkpoint inactivation.

Importantly, p31^{comet} strongly promoted Cdc20 ubiquitination by CP-APC, indicating that it may contribute to disassembly of checkpoint complexes (Fig. 4c). Indeed, when CP-extracts were treated with p31^{comet} alone, some Mad2 dissociated from Cdc20 (Fig. 4d). This dissociation required endogenous UbcH10 in the extracts, as addition of UbcH10^{C114S} along with p31^{comet} blocked Mad2 dissociation. In contrast, when both p31^{comet} and wild-type UbcH10 were added to CP-extracts, Mad2 was completely removed from Cdc20 (Fig. 4d). Similarly, both UbcH10 and p31^{comet} are necessary for the dissociation of checkpoint proteins in intact cells. In cells treated with control short interfering RNAs (siRNAs) and released from nocodazole, Mad2 was readily lost from Cdc20-immunoprecipitates, whereas this dissociation was delayed in cells depleted of UbcH10 or p31^{comet} alone (Fig. 4e). Depletion of both UbcH10 and p31^{comet} had a strongly additive effect, with substantial Mad2 remaining on Cdc20 90 minutes after release from nocodazole (Fig. 4e). Depleted cells were correspondingly delayed in degrading cyclin B1 and progressing through mitosis (Supplementary Fig. 5). Together, these results demonstrate that p31^{comet} lowers the threshold for UbcH10-activity on CP-APC, thereby allowing CP-APC to promote dissociation of checkpoint components and subsequent progression to anaphase.

Our data indicate that APC itself can drive the disassembly of checkpoint complexes and the consequent inactivation of the checkpoint (Fig. 4f). While bound to checkpoint proteins, APC ubiquitinates Cdc20 and possibly other components of the APC–Cdc20–checkpoint protein complex, leading to the release of Mad2 (Fig. 3c, d). While unattached kinetochores are available, released Mad2 may once again capture Cdc20. However, once all kinetochores are properly attached, disassembly of checkpoint complexes predominates. Newly liberated APC could bind and inactivate additional Mad2–Cdc20 complexes leading to a rapid, switch-like transition from metaphase to anaphase. This mechanism is controlled by UbcH10 and p31^{comet}, which promote APC-dependent ubiquitination and by de-ubiquitinating enzymes such as USP44, which antagonize it. The high level of UbcH10 observed in many cancer cells is likely to disrupt this control and promote checkpoint inactivation[19–21]. Because spindle checkpoint activity has been shown to enhance tumour cell killing by anti-mitotic drugs[6,22], we expect agents that lower UbcH10 activity to potentiate the action of these drugs, and thus expand our armamentarium of cancer chemotherapeutics.

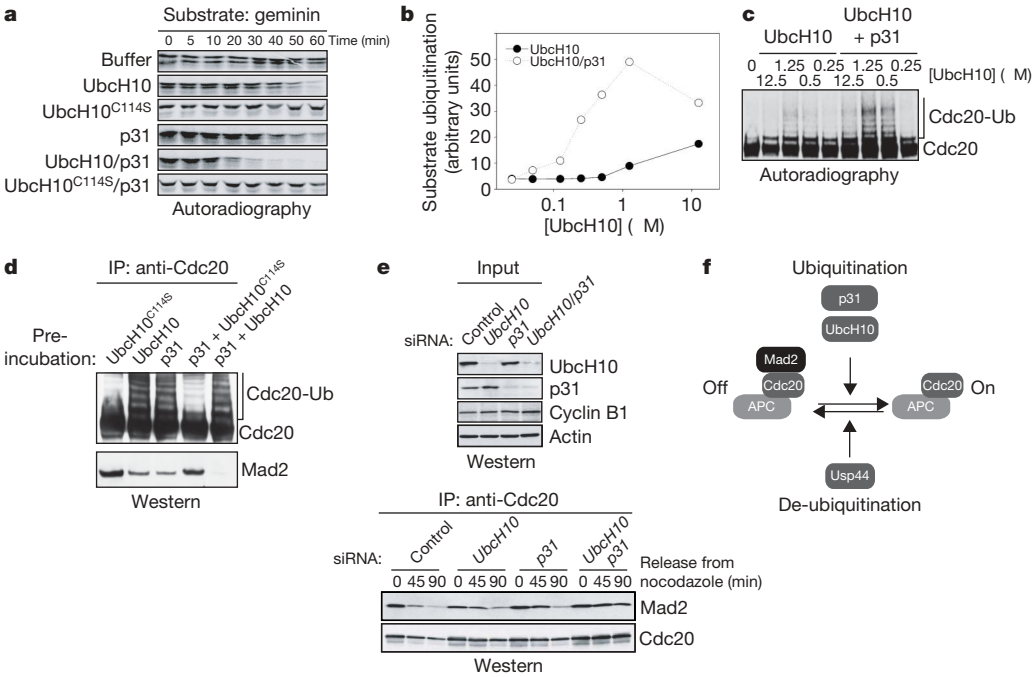


Figure 4 | p31^{comet} lowers the threshold for UbcH10-mediated checkpoint inactivation. **a**, p31^{comet} (p31) and UbcH10 synergistically promote substrate degradation in CP-extracts. CP-extracts were incubated with indicated proteins and degradation of ^{35}S-labelled geminin was detected by autoradiography. **b**, p31^{comet} lowers the threshold for UbcH10-dependent substrate ubiquitination by CP-APC. CP-APC was incubated with UbcH10 alone or in the presence of p31^{comet}. The intensity of ubiquitinated ^{35}S-labelled geminin was quantified by densitometry. **c**, p31^{comet} promotes Cdc20 ubiquitination. ^{35}S-labelled Cdc20 was exchanged into CP-APC, which was then incubated with UbcH10 alone or in the presence of p31^{comet}. Cdc20 ubiquitination was analysed by autoradiography. **d**, UbcH10 and p31^{comet} synergistically promote Mad2 dissociation from Cdc20. CP-extracts were incubated with indicated proteins before Cdc20 immunoprecipitation. Cdc20 and co-purifying Mad2 were detected by western analysis. **e**, UbcH10 and p31^{comet} are both required for efficient inactivation of the checkpoint *in vivo*. HeLa cells were treated with the indicated siRNAs before synchronization. Cdc20 was immunoprecipitated at the indicated times after nocodazole washout. The amount of co-precipitating Mad2 was visualized by western blotting (lower panel). **f**, APC-dependent ubiquitination drives checkpoint complex disassembly and inactivation of the checkpoint. Ubiquitination is promoted by UbcH10 and p31 and is opposed by Usp44.

METHODS

Dissociation assays. For dissociation of checkpoint proteins from purified APC, APC-immunoprecipitates were incubated with 12.5 μM UbcH10 or UbcH10^{C114S} in the presence of ubiquitination components: E1 (Boston Biochem), 20 mM ATP, 1.5 mg ml^{-1} ubiquitin (Boston Biochem), and 10 mM dithiothreitol in UBAB buffer[7] supplemented with 150 mM NaCl and 1 mg ml^{-1} BSA. Reactions were carried out at 23 °C with shaking at 1,000 r.p.m. APC beads were washed extensively and boiled in sample buffer. Bound proteins were examined by SDS–PAGE and western analysis. For dissociation of checkpoint proteins from APC or from Cdc20 in extracts, extracts were incubated with 1.25 μM E2 in the presence of degradation cocktail[7] for 1 h at 23 °C. Extracts were subsequently incubated with anti-Cdc27- or anti-Cdc20-conjugated ProG beads and processed as described above. Alternatively, extracts were centrifuged through gradients of 15–40% sucrose for 18 h at 30,000 r.p.m. in a SW-40 rotor. Fractions were subjected to TCA precipitation, and proteins were subsequently resuspended in sample buffer and analysed by SDS–PAGE and western blotting.

Received 13 December 2006; accepted 8 February 2007.

1. Yu, H. Regulation of APC–Cdc20 by the spindle checkpoint. *Curr. Opin. Cell Biol.* **14,** 706–714 (2002).
2. Musacchio, A. & Hardwick, K. G. The spindle checkpoint: structural insights into dynamic signalling. *Nature Rev. Mol. Cell Biol.* **3,** 731–741 (2002).
3. Cleveland, D. W., Mao, Y. & Sullivan, K. F. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112,** 407–421 (2003).
4. Rieder, C. L. & Maiato, H. Stuck in division or passing through: what happens when cells cannot satisfy the spindle assembly checkpoint. *Dev. Cell* **7,** 637–651 (2004).
5. Stegmeier, F. *et al.* Anaphase initiation is regulated by antagonistic ubiquitination and deubiquitination activities. *Nature* (in the press).
6. Weaver, B. A. & Cleveland, D. W. Decoding the links between mitosis, cancer, and chemotherapy: The mitotic checkpoint, adaptation, and cell death. *Cancer Cell* **8,** 7–12 (2005).
7. Rape, M. & Kirschner, M. W. Autonomous regulation of the anaphase-promoting complex couples mitosis to S-phase entry. *Nature* **432,** 588–595 (2004).
8. Rape, M., Reddy, S. K. & Kirschner, M. W. The processivity of multiubiquitination by the APC determines the order of substrate degradation. *Cell* **124,** 89–103 (2006).
9. Fang, G. Checkpoint protein BubR1 acts synergistically with Mad2 to inhibit anaphase-promoting complex. *Mol. Biol. Cell* **13,** 755–766 (2002).
10. Tang, Z., Bharadwaj, R., Li, B. & Yu, H. Mad2-independent inhibition of APC^{Cdc20} by the mitotic checkpoint protein BubR1. *Dev. Cell* **1,** 227–237 (2001).
11. Hwang, L. H. *et al.* Budding yeast Cdc20: a target of the spindle checkpoint. *Science* **279,** 1041–1044 (1998).
12. Sironi, L. *et al.* Crystal structure of the tetrameric Mad1–Mad2 core complex: implications of a 'safety belt' binding mechanism for the spindle checkpoint. *EMBO J.* **21,** 2496–2506 (2002).
13. Luo, X., Tang, Z., Rizo, J. & Yu, H. The Mad2 spindle checkpoint protein undergoes similar major conformational changes upon binding to either Mad1 or Cdc20. *Mol. Cell* **9,** 59–71 (2002).
14. Luo, X. *et al.* The Mad2 spindle checkpoint protein has two distinct natively folded states. *Nature Struct. Mol. Biol.* **11,** 338–345 (2004).
15. De Antoni, A. *et al.* The Mad1/Mad2 complex as a template for Mad2 activation in the spindle assembly checkpoint. *Curr. Biol.* **15,** 214–225 (2005).
16. Yu, H. Structural activation of Mad2 in the mitotic spindle checkpoint: the two-state Mad2 model versus the Mad2 template model. *J. Cell Biol.* **173,** 153–157 (2006).
17. Xia, G. *et al.* Conformation-specific binding of p31^{comet} antagonizes the function of Mad2 in the spindle checkpoint. *EMBO J.* **23,** 3133–3143 (2004).
18. Mapelli, M. *et al.* Determinants of conformational dimerization of Mad2 and its inhibition by p31^{comet}. *EMBO J.* **25,** 1273–1284 (2006).
19. Wagner, K. W. *et al.* Overexpression, genomic amplification and therapeutic potential of inhibiting the UbcH10 ubiquitin conjugase in human carcinomas of diverse anatomic origin. *Oncogene* **23,** 6621–6629 (2004).
20. Okamoto, Y. *et al.* UbcH10 is the cancer-related E2 ubiquitin-conjugating enzyme. *Cancer Res.* **63,** 4167–4173 (2003).
21. Pallante, P. *et al.* UbcH10 overexpression may represent a marker of anaplastic thyroid carcinomas. *Br. J. Cancer* **93,** 464–471 (2005).
22. Tao, W. *et al.* Induction of apoptosis by an inhibitor of the mitotic kinesin KSP requires both activation of the spindle assembly checkpoint and mitotic slippage. *Cancer Cell* **8,** 49–59 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to H. Yu, F. McKeon, R. King and P. Sorger for generous gifts of reagents, and to F. Stegmeier and S. Elledge for sharing results before publication. We thank members of the Kirschner laboratory, especially P. Jorgensen and M. Springer for helpful discussions. We thank P. Jorgensen, J. Son and J. Schaletzky for critical reading of the manuscript. S.K.R. acknowledges the support of the Medical Scientist Training Program. M.R. was supported by an EMBO long-term fellowship and by a fellowship of the Human Frontiers Science Organization. This work was supported by grants from the National Institutes of Health to M.W.K.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.W.K. (marc@hms.harvard.edu).

LETTERS

Iron meteorite evidence for early formation and catastrophic disruption of protoplanets

Jijin Yang[1], Joseph I. Goldstein[1] & Edward R. D. Scott[2]

In our Solar System, the planets formed by collisional growth from smaller bodies. Planetesimals collided to form Moon-to-Mars-sized protoplanets in the inner Solar System in 0.1–1 Myr, and these collided more energetically to form planets[1]. Insights into the timing and nature of collisions during planetary accretion can be gained from meteorite studies. In particular, iron meteorites offer the best constraints on early stages of planetary accretion because most are remnants of the oldest bodies, which accreted and melted in <1.5 Myr, forming silicate mantles and iron-nickel metallic cores[2–4]. Cooling rates for various groups of iron meteorites suggest that if the irons cooled isothermally in the cores of differentiated bodies, as conventionally assumed, these bodies were 5–200 km in diameter[5,6]. This picture is incompatible, however, with the diverse cooling rates observed within certain groups, most notably the IVA group[7,8], but the large uncertainties associated with the measurements do not preclude it. Here we report cooling rates for group IVA iron meteorites that range from 100 to 6,000 K Myr^{-1}, increasing with decreasing bulk Ni. Improvements in the cooling rate model, smaller error bars, and new data from an independent cooling rate indicator[9] show that the conventional interpretation is no longer viable. Our results require that the IVA meteorites cooled in a 300-km-diameter metallic body that lacked an insulating mantle. This body probably formed ~4,500 Myr ago in a 'hit-and-run' collision between Moon-to-Mars-sized protoplanets[10]. This demonstrates that protoplanets of ~10^3 km size accreted within the first 1.5 Myr, as proposed by theory, and that fragments of these bodies survived as asteroids.

We investigated the group IVA iron meteorites, which come from a single asteroid[5,6], because they present the biggest challenge to existing formation models for iron meteorites. Studies of the formation of the Widmanstätten pattern—the microstructure in which kamacite forms on the close-packed planes of the parent taenite at high temperatures—have concluded that cooling rates for the group IVA iron meteorites decrease by a factor of 10–100 with increasing bulk Ni concentration[7,8], although samples from a mantled core should have indistinguishable cooling rates[11]. In addition, two IVA members contain abundant silicates, which should have floated out of a mantled core before it crystallized. If the observed correlation between cooling rate and Ni concentration is not an artefact of poor sampling (see ref. 12), either the cooling rates (often called metallographic cooling rates) are flawed[5,13], or there is an essential feature in the genesis of IVA iron meteorites ('irons') that has not yet been recognized.

We first determined the cooling rates at 1,000–700 K for ten IVA irons using recent models for nucleation and growth of the Widmanstätten kamacite[14] (see Methods). These range from 100 to 6,600 K Myr^{-1} and decrease monotonically with bulk Ni, as previously reported[7,8], but exceed previous estimates (see Fig. 1, Table 1). Next, we investigated the cloudy zone microstructure at the periphery of lamellae of taenite in seven unreheated and low-shock IVA irons using a transmission electron microscope (TEM) (see Fig. 2), as the dimensions of these microstructures depend on cooling rate at 600–500 K. The apparent absence of any systematic variation of the dimensions of the cloudy zone microstructure with Ni for six IVA irons[9] was interpreted as evidence that the diverse cooling rates determined by the metallographic cooling rate method were flawed[5,13]. We found, however, that the sizes of high-Ni particles in the cloudy zones of IVA irons range from 10 to 29 nm, increasing in size with increasing bulk Ni (Fig. 1, Table 1). As the particle dimensions and metallographic cooling rates are inversely related (Fig. 3), these data are consistent with the cooling rate variation in IVA irons inferred from the Widmanstätten pattern. Although cooling rates cannot be derived directly from cloudy taenite particle sizes because there is no effective model available for spinodal growth, the relative cooling

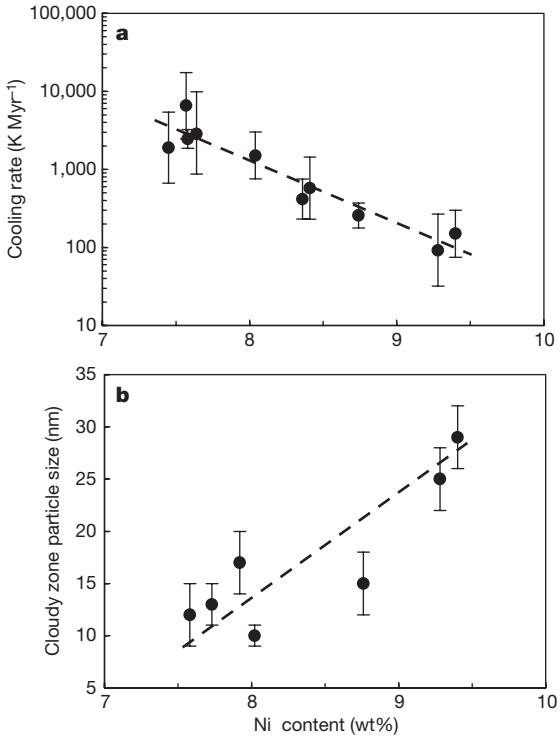


Figure 1 | Dependence of cooling rate and cloudy zone particle size of IVA irons on the bulk meteorite composition. Metallographic cooling rates at 1,000–700 K inferred from kamacite growth modelling (**a**) and the size of cloudy zone particles of low-shock irons (**b**) are plotted as a function of bulk Ni concentration. Error bars: **a**, 2σ uncertainty factors; **b**, ± 1 s.e.m. (see Table 1).

[1]Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst, Massachusetts 01003, USA. [2]Hawaii Institute of Geophysics and Planetology, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA.

Table 1 | Measured and calculated properties of IVA iron meteorites

Meteorite	Ni* (wt%)	Cooling rate (K Myr^{-1})	2σ uncertainty factor†	Shock level (GPa)	Size of CZ particle‡ (nm)
Obernkirchen	7.64	2,900	3.3	40§	
Jamestown	7.45	1,900	2.8	13–40§	
La Grange	7.57	6,600	2.6	13–75§	
Bishop Canyon	7.58	2,500	1.3	<13‖	12 ± 3
San Francisco Mountains	7.73			<13‖	13 ± 2
Bristol	7.92			<13‖	17 ± 3
São João Nepomuceno	8.02			<13‖	10 ± 1
Gibeon	8.04	1,500	2.0	>13‖	
Seneca Township	8.41	580	2.5	13–75§	
Altonah	8.36	420	1.8	13–75§	
Bushman Land	8.76	260	1.4	<13‖	15 ± 3
Steinbach	9.40	150	2.0	<13‖	29 ± 3
Duchesne	9.28	100	2.9	<13§	25 ± 3

* Ref. 13.
† The 2σ uncertainty factors for the cooling rates of individual irons, which range from 1.3 to 3.3, represent a combination of measurement errors and possible errors due to the basic assumptions of the computer model[14], such as chemical homogeneity and uncertainties in the kamacite nucleation temperature. Uncertainties at the 2σ level for the cooling rates of individual irons are 10–20 times smaller than the total range observed within the IVA group.
‡ Data show mean of 4–10 measurements ±1 s.e.m. CZ, cloudy zone.
§ Ref. 29.
‖ Shock level estimated using optical microscopy and SEM.

rates of two meteorites, (CR_1) and (CR_2), can be estimated from the ratio of their respective high-Ni particle sizes (PS_2/PS_1), namely: $(CR_1/CR_2) = (PS_2/PS_1)^n$. The parameter n obtained from Fig. 3 equals 2.4 ± 0.4. Given the threefold variation in IVA particle size (2.9 ± 0.5; Table 1), we estimate that cooling rates at 600–500 K varied by a factor of ~15, decreasing with increasing bulk Ni.

Our data for cloudy zone microstructures in IVA irons are more accurate than previously published data[9]. We used a TEM to directly image the cloudy zone, whereas the earlier study used a scanning electron microscope (SEM) and samples that had been etched to make the cloudy zone visible[9]. In addition, we measured IVA irons that experienced only low shock, <13 GPa, whereas the previous study used some samples that were heavily shocked or showed signs of plastic deformation. Six of the ten IVA irons selected for cooling rate analysis of their Widmanstätten patterns experienced shock pressures >13 GPa and were not used in this study because their cloudy zones were damaged or obliterated. However, shock heating only caused compositional changes at the submicrometre level and did not affect compositions at the micrometre level that were used to derive the cooling rates in Table 1.

Our results from two different techniques show that cooling rates of IVA irons decrease with increasing Ni (7.45–9.4 wt%) by factors of ~50 and ~15 at 1,000–700 K and 600–500 K, respectively (Fig. 1), and cannot be explained by cooling in a mantled core, or by poor sampling of a disrupted and reassembled core[12]. The cooling rates of low-Ni IVA irons at 1,000–700 K are also incompatible with conventional models for asteroidal melting by ^{26}Al. To cool a mantled core at 3,000 K Myr^{-1} would require a body only 4 km in radius, according to thermal models of differentiated asteroids[11]. However, bodies <20 km radius would not have been melted sufficiently by ^{26}Al to form a core[4]. We infer that IVA irons could not have cooled in a mantle-insulated core.

The only plausible solution to the IVA cooling rate enigma is that a metallic body with virtually no silicate mantle was formed by an early catastrophic disruption of a differentiated body and cooled in space so that it had a significant thermal gradient when kamacite nucleated at ~1,000 K. We have tested this concept with thermal modelling of cooling metallic bodies. Figure 4 shows how cooling rate varies with decreasing temperature at different distances from the centre of a metallic body with radius R = 150 km, assuming an initial temperature of 1,750 K. The range of cooling rates of samples between 0.4R and 0.97R decreases from a factor of ~50 during Widmanstätten

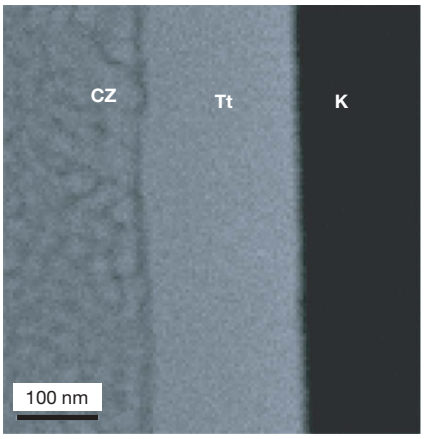


Figure 2 | Microstructure of the cloudy zone in the Steinbach IVA iron. The Ni X-ray map reveals three zones (CZ, cloudy zone; Tt, tetrataenite; K, kamacite). The cloudy zone consists of high-Ni nanoscale particles (light grey), ~50 wt% Ni, that are surrounded by regions of low-Ni phase (dark grey). The size of the high-Ni nanoscale particles in meteorites correlates inversely with cooling rate (Fig. 3). Scale bar, 100 nm.

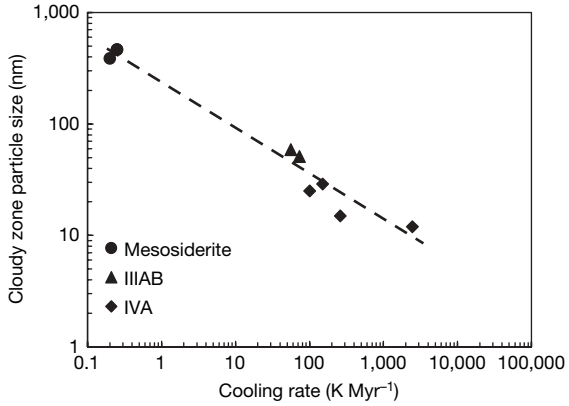


Figure 3 | Variation of the size of the cloudy zone particles in individual mesosiderites and iron meteorites with metallographic cooling rate. This updated plot, unlike the previous plot[9], shows cloudy zone particle sizes for IVA irons that are entirely consistent with the overall trend shown by iron and stony-iron meteorites. Thus, our new cloudy taenite data for IVA irons support the metallographic cooling rate variation within group IVA. Cooling rates are from ref. 30 (mesosiderites), ref. 14 (IIIAB irons) and Table 1 (IVA irons); cloudy zone particle sizes are from ref. 9 (mesosiderites and IIIAB irons) and this study (IVA irons).

growth at 1,000–700 K to a factor of ~15 when cloudy taenite forms at 600–500 K. As these values and the actual cooling rates at 1,000–700 K are very close to those observed for the IVA irons, we conclude that the IVA irons cooled in a metallic body of radius 150 km, and not in an insulated core 3–5 km in radius as previously inferred[6]. Given the inferred 2σ uncertainties in these cooling rates and the likelihood that the IVA irons are reasonably representative of the IVA body[15], the thermal models suggest that the uncertainty in the radius is ~50 km. This metallic body would have been comparable in size to the largest M class asteroid, 6 Psyche. A corollary of this model is that the metallic body crystallized inwards[16] not outwards[5], so that low-Ni irons crystallized on the outside.

We cannot infer from the thermal models whether the metal was initially solid at 1,000–1,300 K or very largely molten above ~1,750 K, as the thermal model is not very sensitive to the initial temperature (see Methods). However, two features suggest that it was initially largely molten. The presence of silicates in five IVA irons has been attributed to a major impact involving silicate melt[12,17]. If the metallic body was initially molten after the impact, the silicates may have been trapped in the metal when it started to crystallize inwards. The extremely low levels of moderately volatile siderophiles in IVA irons (Ge/Ni ratios are 0.01–0.001 times chondritic values) may also result from melting and vaporization during impact[5]. However, the colliding bodies were more likely to have been protoplanets, as proposed for the Moon and Vesta[18], rather than asteroids[5], as the former would have been more efficient at vaporizing metal. These features also suggest that the impact that formed the IVA metallic body occurred before the IVA irons crystallized, ~4,500 Myr ago[6].

Many other meteorite groups have thermal histories that are incompatible with igneous differentiation and cooling in isolated bodies, and require impact disruption at high temperatures. For example, three kinds of silicate-rich, differentiated meteorites have anomalous thermal histories requiring early disruption of molten or partly molten parent bodies: ureilites, mesosiderites and the Shallowater aubrite[19]. In addition, group IIAB, IIIAB and IVB irons have cooling rates that vary by factors of 6–12 (refs 14, 20, 21), which exceed the uncertainties in the cooling rate of individual irons. We infer that the parent bodies of the IIAB, IIIAB and IVB iron meteorites were much larger than previously inferred and cooled with little or no mantle, and that these irons did not cool isothermally inside silicate mantles as widely believed[6,21].

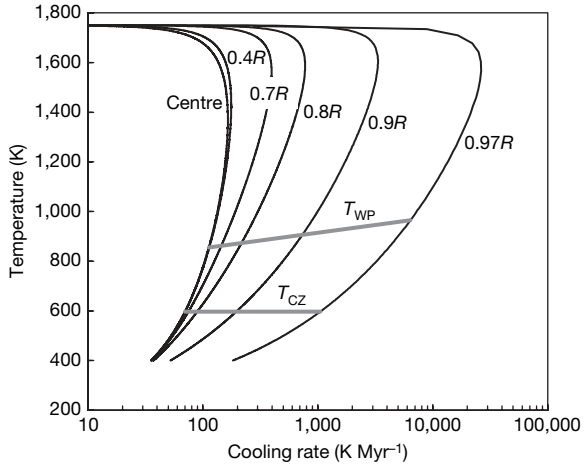


Figure 4 | Variation with radial location and temperature of the cooling rates inside a 150-km-radius solid metallic body exposed to space. We assume that the body was initially at 1,750 K and had a surface temperature of 200 K. T_{WP} and T_{CZ} indicate the respective temperatures at which the Widmanstätten pattern and the cloudy zone began to form. The T_{WP} line is drawn for the IVA irons using the relationship between bulk Ni and cooling rate shown in Fig. 1a.

If the initial mass of the asteroid belt was only a few times the current mass[22], impacts capable of forming a 150-km-radius metallic body would have been exceptionally rare. However, if the initial mass was $\sim 10^3$ times higher and many Moon-sized or larger protoplanets were present[23,24], collision rates would have been greatly enhanced while the differentiated bodies were still hot. As mantles are not efficiently stripped from cores by catastrophic impacts between asteroid-sized bodies[25], the IVA metallic body may have formed as a result of a grazing collision between protoplanets that disrupted the core of a Moon-sized body creating a string of metal-rich bodies, or possibly by tidal disruption alone during a close approach[10]. Opportunities for such collisions would have also have been enhanced if the parent bodies of the irons formed inside 2 AU from the Sun[26].

Given the Hf–W chronological constraints on the origin of iron meteorites and refractory inclusions in chondrites[2,3], we infer that protoplanets with diameters of 10^3 km accreted <1.5 Myr after Solar System formation, as dynamicists suggest[1], and that some were disrupted[10] long before the impacts that exposed the iron meteorites to cosmic rays <1,000 Myr ago. Our study provides the first evidence that fully differentiated bodies were destroyed very early, which is probably an essential factor in understanding the distribution of differentiated asteroids and the rarity of olivine-rich asteroids and meteorites[26].

METHODS

Cooling rate based on the Widmanstätten structure. The Wood method[27] was used to determine the cooling rate by plotting the Ni concentration in the centre of taenite (face-centred cubic Fe-Ni) lamellae—measured using a Cameca SX-50 electron probe microanalyser—against the orientation-corrected taenite half-width measured on the sample surface[14]. A cooling rate code[14] was applied to simulate the diffusion-controlled growth of kamacite (body-centred cubic Fe-Ni) from taenite. The required bulk concentrations of Ni of the meteorites are given in Table 1 and those of P in the Supplementary Table. We also evaluated concerns that the correlation between cooling rate and bulk Ni concentration is an artefact resulting from the use of incorrect nucleation mechanisms or other errors[5,13] (see Supplementary Notes).

Particle size in cloudy zone structure. Techniques used to measure the dimensions of high-Ni particles in IVA cloudy zones are especially critical at the 10 nm level. If etching is used to reveal the microstructure[9], this can affect the apparent size of the microstructure even in a high-resolution SEM. We prepared thin sections of kamacite–taenite interfaces without etching using a dual-beam FEI DB-235 focused ion beam instrument, and examined them in a FEI Tecnai F30ST field emission TEM operated at 300 kV. The dimensions of the high-Ni particles in seven IVA irons that experienced low shock pressure (<13 GPa) and revealed undamaged cloudy zone microstructure were measured (Fig. 1b, Table 1).

Relative cooling rates inferred from the dimensions of cloudy zone microstructures are not a function of the nucleation temperature of the Widmanstätten pattern, as the cloudy zone forms via a spinodal reaction[9]. The spinodal decomposition in meteoritic Fe-Ni metals is a diffusion-controlled reaction, which depends on the cooling rate and composition. For the IVA irons, taenite is P saturated when spinodal decomposition occurs (<600 K) so that the diffusion coefficient of Ni in taenite is the same irrespective of the initial bulk Ni and P content of the IVA meteorite.

Thermal model of an exposed metallic body. As the thermal gradient in a liquid metallic body will be very small until the metal solidifies, we assumed that a spherical solid isothermal Fe-Ni metallic body covered by virtually no silicate mantle at 1,750 K was exposed to space at 200 K following an impact. We used the thermal conductive equation[11] assuming that no heat was generated in the metal during cooling. The Crank-Nicolson approximation and the tridiagonal matrix algorithm[28] were used to describe the partial differential equation for conduction cooling and to solve the difference equation. The physical properties of metal and silicate are from ref. 11.

Our calculations show that the cooling rates at the relevant temperature range (1,000–500 K) for a 150-km-radius metallic body were insensitive to the surface temperature below a plausible maximum of ~300 K, and to the assumed initial temperature over the range 1,750–1,000 K. For example, for an initial temperature of 1,300 K, the range of cooling rates during kamacite growth (1,000–700 K) from <0.4R to 0.96R, where R is the radius of the body, is comparable to the metallographic cooling rates of IVA irons (Table 1). If the initial temperature is 1,750 K (Fig. 4), the corresponding depth range for the IVA irons

becomes $<0.4R$ to $0.97R$. We use a solid body for convenience in thermal modelling, as the cooling rates at 1,000–500 K are also insensitive to the physical condition of the metal immediately after exposure of the metallic body to space, as long as the temperature was >1,000 K.

The effect of radiation from the surface of a 150-km body was also calculated. Heat was effectively radiated away from the surface of the body, so that the surface temperature decreased from 1,750 K to <300 K in only 100 yr. At depths >3 km below the surface, there was no difference between a model that included radiation and a model that assumed a fixed surface temperature of 200 K.

We also calculated the effect of silicate mantle material on the outside of a cooling metallic body of radius 150 km. If the mantle layer is >1 km thick, the cooling rate range across the metallic body is reduced and the thermal model cannot reproduce the fast cooling rates of low-Ni IVA irons near the surface (Fig. 4). A thin silicate mantle could form as a result of late accretion of silicate, incomplete removal, or incomplete separation of silicate that was mixed into molten metal by the impact.

Received 15 September 2006; accepted 26 February 2007.

1. Chambers, J. E. Planetary accretion in the inner Solar System. *Earth Planet. Sci. Lett.* **223,** 241–252 (2004).
2. Kleine, T., Mezger, K., Palme, H., Scherer, E. & Münker, C. Early core formation in asteroids and late accretion of chondrite parent bodies: Evidence from ^{182}Hf-^{182}W in CAIs, metal-rich chondrites, and iron meteorites. *Geochim. Cosmochim. Acta* **69,** 5805–5818 (2005).
3. Scherstén, A., Elliott, T., Hawkesworth, C., Russell, S. & Masarik, J. Hf-W evidence for rapid differentiation of iron meteorite parent bodies. *Earth Planet. Sci. Lett.* **241,** 530–542 (2006).
4. Hevey, P. J. & Sanders, I. S. A model for planetesimal meltdown by ^{26}Al, and its implications for meteorite parent bodies. *Meteorit. Planet. Sci.* **41,** 95–106 (2006).
5. Wasson, J. T., Matsunami, Y. & Rubin, A. E. Silica and pyroxene in IVA irons; possible formation of the IVA magma by impact melting and reduction of L–LL-chondrite materials followed by crystallization and cooling. *Geochim. Cosmochim. Acta* **70,** 3149–3172 (2006).
6. Chabot, N. L. & Haack, H. in *Meteorites and the Early Solar System II* (eds Lauretta, D. S. & McSween, H. Y.) 747–771 (Univ. Arizona Press, Tucson, 2006).
7. Moren, A. E. & Goldstein, J. I. Cooling rates of group IVA iron meteorites determined from a ternary Fe-Ni-P model. *Earth Planet. Sci. Lett.* **43,** 82–96 (1979).
8. Rasmussen, K. L., Ulff-Moller, F. & Haack, H. The thermal evolution of IVA iron meteorites: Evidence from metallographic cooling rates. *Geochim. Cosmochim. Acta* **59,** 3049–3059 (1995).
9. Yang, C.-W., Williams, D. B. & Goldstein, J. I. A new empirical cooling rate indicator for meteorites based on the size of the cloudy zone of the metallic phase. *Meteorit. Planet. Sci.* **32,** 423–429 (1997).
10. Asphaug, E., Agnor, C. B. & Williams, Q. Hit-and-run planetary collisions. *Nature* **439,** 155–160 (2006).
11. Haack, H., Rasmussen, K. L. & Warren, P. H. Effects of regolith/megaregolith insulation on the cooling histories of differentiated asteroids. *J. Geophys. Res.* **95,** 5111–5124 (1990).
12. Haack, H., Scott, E. R. D., Love, S. G., Brearley, A. J. & McCoy, T. J. Thermal histories of IVA stony-iron and iron meteorites: Evidence for asteroid fragmentation and reaccretion. *Geochim. Cosmochim. Acta* **60,** 3103–3113 (1996).
13. Wasson, J. T. & Richardson, J. W. Fractionation trends among IVA iron meteorites: contrast with IIIAB trends. *Geochim. Cosmochim. Acta* **65,** 951–970 (2001).
14. Yang, J. & Goldstein, J. I. Metallographic cooling rates of the IIIAB iron meteorites. *Geochim. Cosmochim. Acta* **70,** 3197–3215 (2006).
15. Scott, E. R. D., Haack, H. & McCoy, T. J. Core crystallization and silicate-metal mixing in the parent body of the IVA iron and stony-iron meteorites. *Geochim. Cosmochim. Acta* **60,** 1615–1631 (1996).
16. Haack, H. & Scott, E. R. D. 1992. Asteroid core crystallization by inward dendritic growth. *J. Geophys. Res.* **97,** 14727–14734 (1992).
17. Ruzicka, A. & Hutson, M. Differentiation and evolution of the IVA meteorite parent body: clues from pyroxene geochemistry in the Steinbach stony-iron meteorite. *Meteorit. Planet. Sci.* **41,** 1959–1987 (2006).
18. Halliday, A. N. & Kleine, T. in *Meteorites and the Early Solar System II* (eds Lauretta, D. S. & McSween, H. Y.) 775–801 (Univ. Arizona Press, Tucson, 2006).
19. Keil, K., Haack, H. & Scott, E. R. D. Catastrophic fragmentation of asteroids: evidence from meteorites. *Planet. Space Sci.* **42,** 1109–1122 (1994).
20. Randich, E. & Goldstein, J. I. Cooling rates of seven hexahedrites. *Geochim. Cosmochim. Acta* **42,** 221–233 (1978).
21. Rasmussen, K. L. Cooling rates and parent bodies of iron meteorites from group IIICD, IAB, and IVB. *Phys. Scripta* **39,** 410–416 (1989).
22. Davis, D. R., Durda, D. D., Marzari, F., Bagatin, A. C. & Gil-Hutton, R. G. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 545–558 (Univ. Arizona Press, Tucson, 2002).
23. Petit, J.-M., Chambers, J., Franklin, F. & Nagasawa, M. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 711–723 (Univ. Arizona Press, Tucson, 2002).
24. Bottke, W. F. *et al.* Linking the collisional history of the main asteroid belt to its dynamical extinction and depletion. *Icarus* **179,** 63–94 (2005).
25. Scott, E. R. D., Haack, H. & Love, S. G. Formation of mesosiderites by fragmentation and reaccretion of a large differentiated asteroid. *Meteorit. Planet. Sci.* **36,** 869–881 (2001).
26. Bottke, W. F., Nesvorny, D., Grimm, R. E., Morbidelli, A. & O'Brien, D. P. Iron meteorites as remnants of planetesimals formed in the terrestrial planet region. *Nature* **439,** 821–824 (2006).
27. Wood, J. A. The cooling rates and parent bodies of several iron meteorites. *Icarus* **3,** 429–459 (1964).
28. Von Rosenberg, D. U. *Methods for the Numerical Solution of Partial Differential Equations* (American Elsevier Publishing Company, New York, 1969).
29. Jain, A. V. & Lipschutz, M. E. On preferred disorder and shock history of chemical group IVA meteorites. *Geochim. Cosmochim. Acta* **34,** 883–892 (1970).
30. Hopfe, W. D. & Goldstein, J. I. The metallographic cooling rate method revised: Application to iron meteorites and mesosiderites. *Meteorit. Planet. Sci.* **36,** 135–154 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support from the NASA Cosmochemistry programme is acknowledged. We thank T. J. McCoy, L. Welzenbach, D. S. Ebel and J. Boesenberg for providing the meteorite samples; M. J. Jercinovic, J. R. Michael and P. Kotula for assistance and advice in obtaining electron microprobe and microscopy data; G. J. Taylor, E. Asphaug, W. F. Bottke and A. Ruzicka for discussions; and H. Haack, J. T. Wasson and J. A. Wood for critically reading the manuscript.

Author Contributions J.Y., J.I.G. and E.R.D.S. contributed equally to this work. J.Y. determined the metallographic cooling rates and established the thermal model. J.I.G. measured the cloudy zone particle size. E.R.D.S. provided the planetary science perspective. All authors discussed the results, wrote portions of the paper and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.Y. (jiyang@ecs.umass.edu).

Vol 446|19 April 2007|doi:10.1038/nature05739

LETTERS

A new perceptual illusion reveals mechanisms of sensory decoding

Mehrdad Jazayeri[1] & J. Anthony Movshon[1]

Perceptual illusions are usually thought to arise from the way sensory signals are encoded by the brain, and indeed are often used to infer the mechanisms of sensory encoding[1]. But perceptual illusions might also result from the way the brain decodes sensory information[2], reflecting the strategies that optimize performance in particular tasks. In a fine discrimination task, the most accurate information comes from neurons tuned away from the discrimination boundary[3,4], and observers seem to use signals from these 'displaced' neurons to optimize their performance[5,6,7]. We wondered whether using signals from these neurons might also bias perception. In a fine direction discrimination task using moving random-dot stimuli, we found that observers' perception of the direction of motion is indeed biased away from the boundary. This misperception can be accurately described by a decoding model that preferentially weights signals from neurons whose responses best discriminate those directions. In a coarse discrimination task, to which a different decoding rule applies[4], the same stimulus is not misperceived, suggesting that the illusion is a direct consequence of the decoding strategy that observers use to make fine perceptual judgments. The subjective experience of motion is therefore not mediated directly by the responses of sensory neurons, but is only developed after the responses of these neurons are decoded.

Subjects viewed a field of moving dots within a circular aperture around a fixation point for 1 s and reported whether the direction of motion was clockwise (CW) or counterclockwise (CCW) of a decision boundary indicated by a bar outside the edge of the dot-field (Fig. 1a). On each trial, the boundary took a random position around the dot field, and a percentage of dots (3%, 6% or 12%) moved coherently in a randomly chosen direction within 22 degrees of the boundary; the other dots moved randomly. After each trial, subjects pressed one of two keys to indicate their choice (CW or CCW). For 70% of trials, they were given feedback. On the remaining

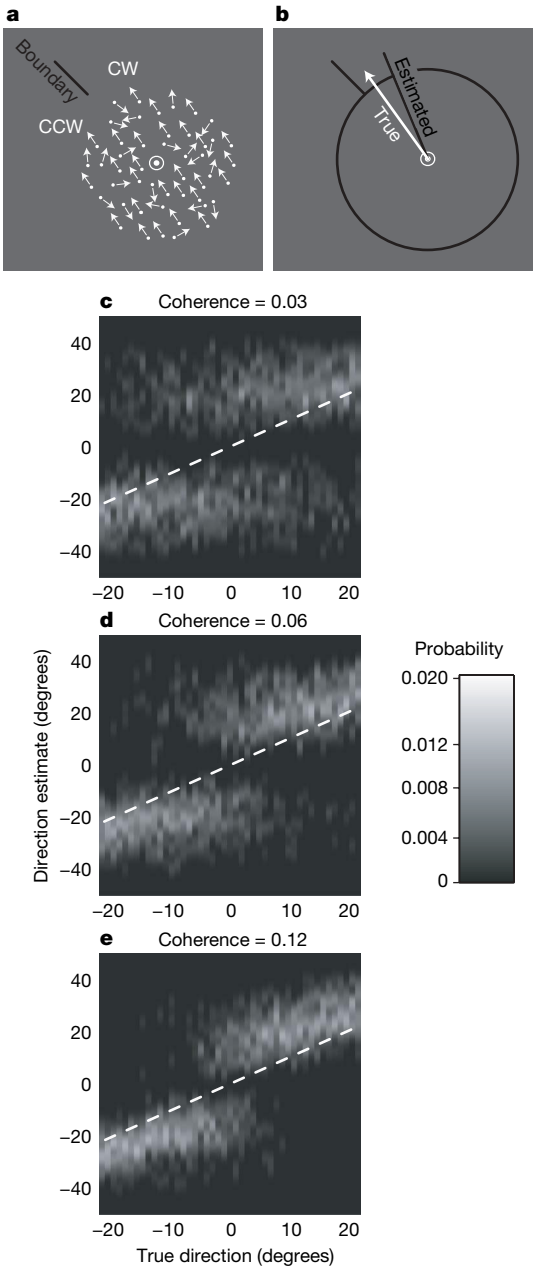


Figure 1 | The combined discrimination–estimation experiment. **a**, The discrimination phase. Subjects viewed a field of moving random dots and indicated whether its direction was clockwise (CW) or counter-clockwise (CCW) with respect to an indicated discrimination boundary that varied randomly from trial to trial. **b**, The estimation phase. On an unpredictable 30% of trials, after discriminating the direction of motion, subjects reported their estimate of the direction of motion by extending a dark line from the centre of the display with the computer mouse. **c**–**e**, Image maps representing the distribution of estimation responses for one subject at the three coherence levels. Each column of each plot represents the distribution of estimates for a particular true direction of motion, using a nonlinear lightness scale for probability (right). The observed values have been smoothed parallel to the ordinate with a gaussian (s.d. = 2 degrees) for clarity. The black dashed line is the locus of veridical estimates. Responses in the top-left and bottom-right quadrants of each map correspond to error trials, whereas those in the top-right and bottom-left quadrants were for correct trials; the discrimination and estimation responses were concordant throughout. Judgement errors (top-left and bottom-right quadrants) decrease with increasing coherence and as direction becomes more different from the discrimination boundary.

[1]Center for Neural Science, New York University, 4 Washington Place, New York, New York 10003, USA.

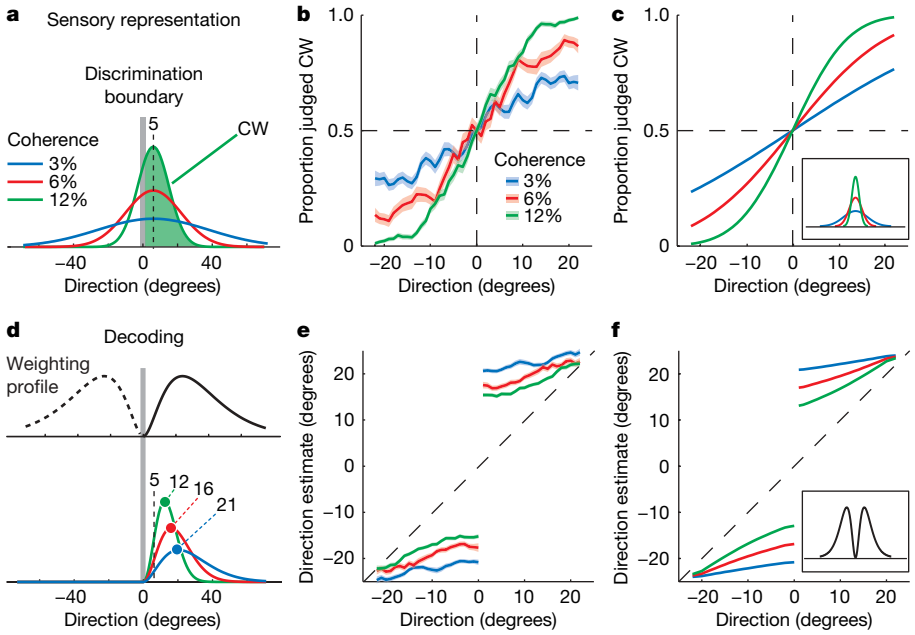


Figure 2 | Discrimination and estimation responses. **a**, The sensory representations evoked by a dot field moving 5° CW vary from one trial to the next. The plot is a cartoon of the distribution of these noise-perturbed sensory representations for different coherences. The distribution becomes more variable with weaker signals. The proportion of CW judgements is the area under the CW part of this distribution, shown for 12% coherence by the shaded green area. **b**, Proportion of CW judgements (thick lines) and their standard errors (shading) as a function of direction of motion for all coherence values for one subject. The CW and CCW portions of the data from Fig. 1b have been pooled and smoothed with a 3 degree boxcar filter. **c**, Fits to the discrimination data in **b**, using the model drawn in **a**; the inset shows the inferred sensory representations. **d**, The decoding model. The sensory representations from **a** are multiplied by a displaced weighting profile that is optimal for discriminating CW from CCW alternatives. As a result, the peaks of the distributions shift away from the boundary. The plot shows schematically how this model predicts larger shifts for lower coherence values—the peaks for coherences of 3%, 6% and 12% fall at 12, 16 and 21 degrees respectively, even though the peak of the underlying sensory representation (from **a**) remains at 5 degrees. **e**, Subjective estimates as a function of direction of motion for trials on which motion direction was correctly discriminated. The CW and CCW portions of the data (Fig. 1d) have been pooled and smoothed with a 3 degree boxcar filter. **f**, The model fits for the subjective estimates after estimating and applying the single weighting profile that best matches the data.

30%, feedback was withheld and subjects estimated the direction of motion they had seen by aligning a bar extending from the fixation point to the direction of their estimate (Fig. 1b).

For all subjects, discrimination performance was lawfully related to motion coherence and direction: performance improved for higher coherences and for directions of motion farther away from the boundary, and there was no systematic bias in the choice behaviour (Fig. 1c–e). However, when subjects were asked to report the direction of motion, their estimates deviated from the direction of motion in the stimulus, and were biased in register with their discrimination choice (Fig. 1c–e). The magnitude of these deviations depended on both the coherence and the direction of motion, being

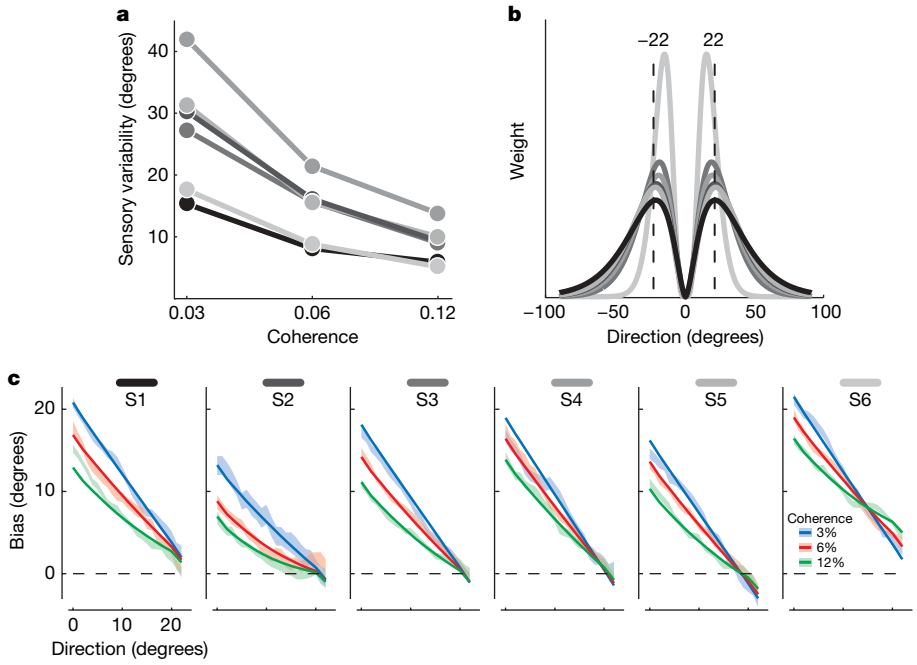


Figure 3 | Summary data for all six subjects. **a**, The variability of the sensory representations as a function of motion coherence for all subjects (computed as the standard deviation of the gaussian fits, for example, Fig. 2c, inset) are shown with different shades of grey (S1 to S6 in **c**). **b**, Recovered weighting functions for all six subjects. The dotted lines delimit the range of directions of motion (−22 to 22 degrees) that were used in the experiment. **c**, The mean bias (the difference between the true and estimated directions) ± one standard error (shading) and the model fits (thick line) for all subjects and all coherence values.

larger for more uncertain conditions when either coherence was low or the direction was close to the boundary—the conditions in which discrimination performance was worst.

The sensory representation evoked by the random dot stimulus is perturbed by noise[8,9,10], and we therefore expect it to be more variable from trial to trial for weak motion signals than for stronger ones (Fig. 2a). How well observers discriminate the alternatives depends on the strength of the motion signal and its direction with respect to the boundary. Assuming that the variability in the sensory representation can be described by a gaussian, we fitted the discrimination performance of each subject (Fig. 2b) with a cumulative gaussian to estimate the spread of the sensory representation for each level of coherence. As expected, the variance of this distribution decreased with increasing coherence (Fig. 2c, inset).

This formulation accounts simply and well for discrimination behaviour (Fig. 2c), but it does not explain why the subjective estimates deviate from the true direction of motion in the stimulus. To understand what causes the perceptual biases, we considered the events that lead to the subjective estimates of the direction of motion. On each trial, before reporting their estimate, observers make a fine perceptual judgement. To do so, they have to transform the sensory responses into a binary decision (CW or CCW). Because subjects did not receive feedback on their subjective estimates, they could only adjust their decoding strategy for the discrimination part of the combined discrimination–estimation task where they did receive feedback. As shown both in theory[3,4] and experiment[6,7], in a fine discrimination task like ours, neurons with direction preferences moderately shifted to the sides of the boundary make the largest contribution, whereas neurons tuned to directions either near or very remote from the boundary are less important. Therefore, to decode the activity of sensory neurons efficiently, the brain must pool their responses with a weighting profile that has maxima moderately shifted to the sides of the boundary[4] (Fig. 2d, top panel).

If the pattern of direction estimates is explained by such a displaced profile, there should be a weighting function which, when applied to the sensory representation of different stimuli, predicts the corresponding estimates. We computed the product of this weighting profile (Fig. 2d, top panel) with the sensory representation of the stimulus estimated from the discrimination performance (Fig. 2a), and took the peak as the direction estimate (Fig. 2d, bottom panel). For each observer, we fitted the weighting profile that, when combined with that observer's discrimination performance, best predicted the pattern of direction estimates. Remarkably, combining the sensory representation with a single weighting profile (Fig. 2f, inset) accurately captured the observed estimates for all coherence levels and all directions of motion (Fig. 2f). Though observers varied in the accuracy of their sensory representations (Fig. 3a), the inferred weighting functions were similar for all six (Fig. 3b), and the resulting model accurately predicted the estimation bias (the difference between the true and estimated directions) for all six (Fig. 3c).

The misperception of motion can be economically attributed to the decoding strategy that observers adopt to optimize fine perceptual judgements, but other interpretations are possible. For example, the misperception might reflect a change in the sensory representation evoked by the stimulus, and not the way it is decoded. To test this idea, we ran a second experiment that differed from the first only in that the fine discrimination was replaced by coarse discrimination. On every trial, we presented motion in a randomly chosen direction within 22 degrees of a bar presented in the periphery (previously used for discrimination boundary), or within 22 degrees of the direction opposite the bar. Subjects discriminated whether the direction of motion was towards or away from the bar and as before, on a subset of trials reported their estimate of the direction of motion. As shown in theory[4] and experiment[11], the most accurate information now comes from neurons tuned to the two alternatives. Therefore, the bias should, if anything, change from repulsion to attraction. This is exactly the pattern of responses we observed (Fig. 4a–c). The illusion thus depends entirely on the subject's task—it occurs during fine discrimination but not during coarse discrimination (see Supplementary Information for a more detailed discussion). Changes in the sensory representation therefore cannot explain the effect.

One other possibility is that observers did not 'truly' misperceive the motion, but when uncertain about its direction, adopted a biased response strategy to ensure that they would not disagree with their immediately preceding discrimination choice. Simple models of response bias that only depend on the preceding choices can easily be discarded because they cannot account for the systematic relationship between the subjective estimates and the strength and direction of motion (Figs 1c–e, 2e). As detailed in the Supplementary Discussion, biased response strategies that are rich enough to account for our data have to incorporate computations effectively the same as those that our decoding model employs, applying a weighting profile to the sensory representation. Should we then view these computations in the framework of sensory decoding or complex response bias? In our decoding model, all computations serve a well-grounded function that can be inferred from theoretical and experimental observations of fine discrimination[3,4,6,7]. This model also has the

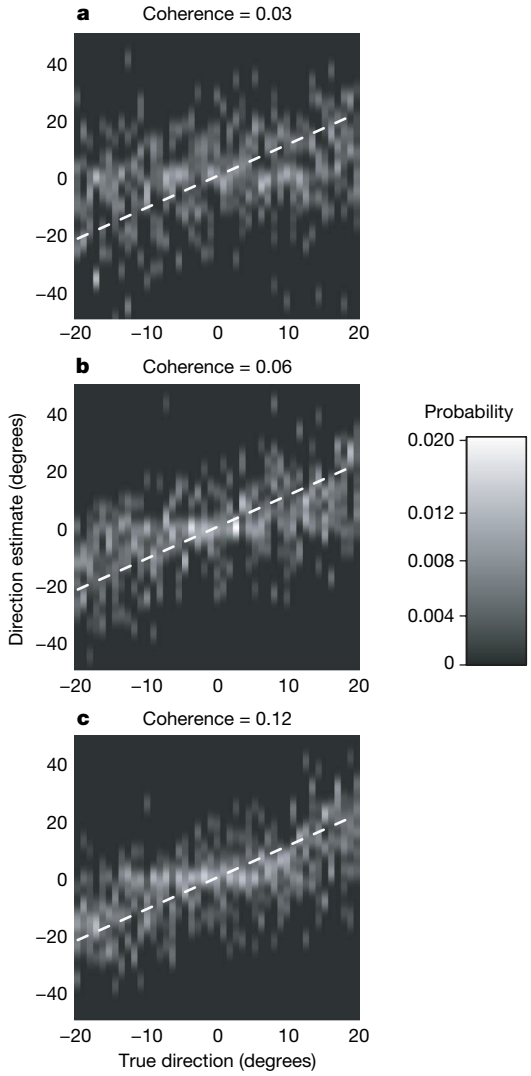


Figure 4 | Subjective estimates in a coarse direction discrimination task. **a–c**, The estimates are represented as in Fig. 1c–e. In the discrimination phase, the subject's task was to indicate whether the direction of motion in the random-dot was towards or away from a peripheral visual cue (the same as the discrimination boundary in Fig. 1b). The distributions for correct discriminations towards and away from the visual cue are pooled (separate plots for the two conditions are shown in Supplementary Fig. 3).

virtue of simplicity, because it accounts for observers' subjective reports using only the machinery that accounts for their objective discrimination performance. Response bias models, on the other hand, postulate two unrelated mechanisms, one to account for discrimination and the other for perceptual reports.

Bias arising from a decoding strategy may also explain some other perceptual distortions, such as those related to repulsion away from the cardinal axes[12] or from other discrimination boundaries[13]. We believe that this 'reference repulsion' phenomenon arises when subjects implicitly discriminate stimulus features against available internal or external references, such as a cardinal direction or the boundary marker in our experiment. This causes their perception of those features to be shifted away from the reference by the mechanism we have described. In other words, these incidences of misperception reveal the optimality of the system—not in perceiving, but in decoding sensory signals to make fine perceptual judgments.

Because the misperception does not seem to reflect the sensory responses to the direction of motion, the subjective experience of motion must be mediated by the machinery that decodes the responses of motion-sensitive neurons. We have argued elsewhere that areas downstream of sensory representations recode sensory responses into sensory likelihoods[4], and the discrimination model used here is derived directly from that representation. Our results therefore suggest that the subjective experience of sensory events arises from the representation of sensory likelihoods, and not directly from the responses of sensory neuron populations.

METHODS

Eight subjects aged 19 to 35 participated in this study after giving informed consent. All had normal or corrected-to-normal vision, and all except one were naive to the purpose of the experiment. Subjects viewed all stimuli binocularly from a distance of 71 cm on an Eizo T960 monitor driven by a Macintosh G5 computer at a refresh rate of 120 Hz in a dark, quiet room.

In the main experiment, in which six of the subjects participated, each trial began with the presentation of a fixation point along with a dark bar in the periphery representing the discrimination boundary for the subsequent motion discrimination (Fig. 1a). After 0.5 s, the motion stimulus was presented for 1 s. Subjects were asked to keep fixation during the presentation of the motion stimulus. After the motion stimulus was extinguished, subjects pressed one of two keys to report whether the direction of motion was CW or CCW with respect to the boundary and received distinct auditory feedback for correct and incorrect judgements. On approximately 30% of trials chosen at random, feedback was withheld and a circular ring was presented as a cue for the subject to report the direction of motion in the stimulus (Fig. 1b). The subject reported the estimate by using a mouse to extend a dark bar from the fixation point in the direction of their estimate and terminated the trial by pressing a key. Subjects were asked to estimate accurately but did not receive feedback. The discrimination boundary and the fixation point persisted throughout the trial. Trials were separated with a 1.5 s inter-trial interval during which the screen was blank.

For the main experiment, subjects had ample time to practice and master the task contingencies. Data for the main experiment were collected only after the discrimination thresholds stabilized (changed less than 10% across consecutive sessions). After this period, subjects completed roughly 8,000 trials in 10–12 sessions, each lasting approximately 45 min.

The remaining two subjects participated in the control coarse discrimination experiment. This experiment differed from the main experiment in that motion was within 22 degrees either towards or opposite to the peripherally presented bar (that is, the discrimination boundary in the main experiment), and during the discrimination stage, subjects had to report whether motion was towards or away from the bar. On approximately 30% to 50% of the trials chosen at random, the feedback was withheld and subjects were asked to report the perceived direction of motion (same procedure as in the main experiment).

All stimuli were presented on a dark grey background of 11 cd m^{-2}. The fixation point was a central circular white point subtending 0.5 degrees with a luminance of 77 cd m^{-2}. A gap of 1 degree between the fixation point and the motion stimulus helped subjects maintain fixation. The discrimination boundary was a black bar 0.5 degrees by 0.15 degrees, 3.5 degrees from fixation. The motion stimulus was a field of dots (each 0.12 degrees in diameter with a luminance of 77 cd m^{-2}) contained within a 5-degree circular aperture centred on the fixation point (Fig. 1). On successive video frames, some dots moved coherently in a designated direction at a speed of 4 deg s^{-1}, and the others were replotted at random locations within the aperture. On each trial, the percentage of coherently moving dots (coherence) was randomly chosen to be 3, 6 or 12%, and their direction was randomly set to a direction within 22 degrees of the discrimination boundary; in the second experiment, half the trials presented motion within 22 degrees of a direction 180 degrees away from the boundary. The dots had an average density of 40 dots deg^{-2} s^{-1}. The presentation of a black circular ring with a radius of 3.3 degrees around the fixation cued the subjects to report their estimate, which they did by moving the mouse to extend and align a black bar of width 0.15 deg to the direction of their estimate.

We modelled the sensory representation with a gaussian probability density function centred at the true direction of motion. The variance of this distribution for each subject was estimated by fitting a cumulative gaussian to his/her discrimination performance to maximize the likelihood of observing the subjects' choices for each level of motion coherence. To predict the direction estimates, we multiplied this sensory representation by a weighting function, and took the peak of the result. An additional additive constant accounted for any motor bias independent of sensory evidence. We chose a gamma probability density function as a convenient parametric form for the weighting profile. For each subject, we obtained parametric fits by minimizing the squared error of the model's prediction for the observed mean direction estimates for that subject. The gamma distribution provided a good fit for our data, but our conclusions do not depend on the exact form of the weighting profile. This simple procedure of first finding the sensory likelihoods from discrimination data, and then computing the weighting profile from the estimation data, crystallizes the contrast between encoding and decoding in our model. In detail, however, it neglects the subtle effect of the weighting profile on discrimination behaviour. In Supplementary Methods, we detail a complete model that simultaneously accounts for both the discrimination choices and the direction estimates in a single step (Supplementary Fig. 1), and present the fits for the correct as well as error trials (Supplementary Fig. 2).

Received 7 November 2006; accepted 9 March 2007.
Published online 4 April 2007.

1. Eagleman, D. M. Visual illusions and neurobiology. *Nature Rev. Neurosci.* **2,** 920–926 (2001).
2. Gregory, R. L. *Eye and Brain: The Psychology of Seeing* 5th edn (Oxford Univ. Press, 1997).
3. Seung, H. S. & Sompolinsky, H. Simple models for reading neuronal population codes. *Proc. Natl Acad. Sci. USA* **90,** 10749–10753 (1993).
4. Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nature Neurosci.* **9,** 690–696 (2006).
5. Patterson, R. D. Auditory filter shapes derived with noise stimuli. *J. Acoust. Soc. Am.* **59,** 640–654 (1976).
6. Regan, D. & Beverley, K. I. Postadaptation orientation discrimination. *J. Opt. Soc. Am. A* **2,** 147–155 (1985).
7. Hol, K. & Treue, S. Different populations of neurons contribute to the detection and discrimination of visual motion. *Vision Res.* **41,** 685–689 (2001).
8. Parker, A. J. & Newsome, W. T. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21,** 227–277 (1998).
9. Dean, A. F. The variability of discharge of simple cells in the cat striate cortex. *Exp. Brain Res.* **44,** 437–440 (1981).
10. Tolhurst, D. J., Movshon, J. A. & Dean, A. F. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* **23,** 775–785 (1983).
11. Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12,** 4745–4765 (1992).
12. Huttenlocher, J., Hedges, L. V. & Duncan, S. Categories and particulars: Prototype effects in estimating spatial location. *Psychol. Rev.* **98,** 352–376 (1991).
13. Rauber, H. & Treue, S. Reference repulsion when judging the direction of visual motion. *Perception* **27,** 393–402 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a research grant from the NIH. We are grateful to B. Lau, E. Simoncelli, D. Heeger, M. Landy and N. Graham for advice and discussion.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.J. (mjaz@cns.nyu.edu).

CORRIGENDUM

doi:10.1038/nature05790

Postnatal isl1^{+} cardioblasts enter fully differentiated cardiomyocyte lineages

K.-L. Laugwitz, A. Moretti, J. Lam, P. Gruber, Y. Chen, S. Woodard, L.-Z. Lin, C.-L. Cai, M. M. Lu, M. Reth, O. Platoshyn, J. X.-J. Yuan, S. Evans & K. R. Chien

Nature 433, 647–653 (2005)

It has been drawn to *Nature*'s attention that K.R.C., S.E., C.-L.C., A.M. and K.-L.L. filed a patent application relevant to this work (patent number WO 2004/070013) in 2004, which should therefore have been declared as a competing financial interest.

CORRIGENDUM

doi:10.1038/nature05791

Global warming and climate forcing by recent albedo changes on Mars

Lori K. Fenton, Paul E. Geissler & Robert M. Haberle

Nature 446, 646–649 (2007)

It has been brought to our attention that there was some ambiguity in the wording of the first boldface paragraph of this Letter regarding our prediction of the change in temperature on Mars that would have resulted from changes in albedo. As is clear from Table 1, we predict a total change of 0.65 degrees Celsius over the period from the 1970s to the 1990s, rather than 0.65 degrees Celsius per year.

Published online 5 April 2007.